

Improving clinical care for newborns in Kenyan hospitals

Dissertation (PhD)

Newton Opiyo

Institute of Health and Society

Faculty of Medicine, University of Oslo

2012

© **Newton Opiyo, 2013**

*Series of dissertations submitted to the
Faculty of Medicine, University of Oslo
No. 1494*

ISBN 978-82-8264-578-2

All rights reserved. No part of this publication may be
reproduced or transmitted, in any form or by any means, without permission.

Cover: Inger Sandved Anfinsen.
Printed in Norway: AIT Oslo AS.

Produced in co-operation with Akademika publishing.
The thesis is produced by Akademika publishing merely in connection with the
thesis defence. Kindly direct all inquiries regarding the thesis to the copyright
holder or the unit which grants the doctorate.

Contents

Acknowledgement	6
Abbreviations	7
List of papers	8
Additional papers related to the dissertation.....	9
Summary	10
Introduction	14
Clinical diagnosis of severe disease in young infants (Paper I)	15
Implementing recommendations into routine health worker practice – the value of training in emergency care.....	15
Newborn resuscitation training (Paper II)	15
In-service neonatal and paediatric emergency care courses (Paper III).....	16
Exploring methods to support development of evidence-informed, contextually appropriate national guidelines for neonatal care	17
Summarising and presenting evidence (Paper IV)	17
The evidence translation process (Paper V)	17
Aims and objectives	19
Methods	20
What clinical signs best identify severe illness in young infants aged 0 to 59 days in developing countries? A systematic review (Paper I).....	20
Search strategy and selection criteria.....	20
Assessment of quality of evidence	20
Effect of newborn resuscitation training on health worker practices in Pumwani Hospital, Kenya: a randomised controlled trial (Paper II).....	21
Study setting	21
Participants and randomisation	21
Intervention	21
Outcome measures	21
Data collection	21
Data analysis	22
In-service training for health professionals to improve care of the seriously ill newborn or child in low- and middle-income countries: a Cochrane review (Paper III)	22

Search strategy and selection criteria.....	22
Assessment of quality of evidence	22
Comparison of alternative evidence summary and presentation formats in clinical guideline development: a mixed-method evaluation (Paper IV)	23
Study design and participants	23
Interventions.....	23
Randomisation	23
Outcome measures.....	24
Data collection	24
Analysis	24
Exploring the evidence translation process during the development of evidence-based guidelines for newborn and paediatric care in Kenya: an observational study (Paper V)	25
Study design and participants.....	25
Child Health Evidence Week	25
Data collection	25
Data analysis	25
Ethics.....	27
Papers I and III.....	27
Paper II	27
Papers IV and V	27
Results.....	28
Paper I	28
Paper II	28
Paper III	29
Paper IV	29
Quantitative findings	29
Results from the interview study.....	30
Views on different formats for presenting systematic review evidence (see Panel 1, Paper IV, for illustrative quotes)	30
Panelists' experiences with appraisal and use of research evidence (see Panel 2, Paper IV, for illustrative quotes).....	30
Paper V	31
Discussion	33

Internal validity	33
Causation and bias minimization	33
Sampling.....	34
External validity.....	35
Discussion of main results.....	37
Clinical diagnosis of severe disease in young infants.....	37
Newborn resuscitation training	38
In-service neonatal and paediatric emergency care courses	39
Evidence summary formats in guideline development	40
Translation of evidence into recommendations.....	41
Intergration of findings and implications for future research	42
Conclusion.....	45
References	46
Papers I-V	51

Acknowledgement

I would like to acknowledge the following people and institutions for their support in this project.

First, special thanks to my supervisors, Atle Fretheim, Sasha Shepperd and Mike English, for their guidance and advice throughout my training. Thanks also to Nyokabi Ruth Musila for valuable comments and contribution in the guideline development studies. I would also like to thank Greg Fegan and Elizabeth Allen for their statistical support. Sincere thanks to all the co-authors for their inputs in the five papers presented in this thesis.

I would like to thank all the nurses at Pumwani Maternity Hospital, resuscitation observers and Newborn Life Support instructors for their help with the newborn resuscitation study. I also thank the mothers in Pumwani Maternity Hospital who agreed to allow observers to be present at the delivery of their babies. I am grateful to all the participants in the 'Child Health Evidence Week' for generously sharing their knowledge and experiences in the guideline development study.

I would like to acknowledge the following organisations for their financial support: The Laerdal Foundation (Stavanger, Norway) for funding the newborn resuscitation study and KEMRI-Wellcome Trust Research Programme for funding the guideline development workshop and my training in Norway. I also present my thanks to the Institute of Health and Society at the University of Oslo for hosting me and facilitating my training in Norway. Special thanks to Ragnhild Beyrer and Terese Eriksen for their administrative support.

Finally, thanks to my family for timeless support and encouragement.

Abbreviations

APGAR	Activity, Pulse, Grimace, Appearance and Respiration
ARI	Acute Respiratory Infection
CDDs	Control of Diarrheal Diseases
CENTRAL	Cochrane Central Register of Controlled Trials
CINAHL	Cumulative Index of Nursing and Allied Health
ENC	Essential Newborn Care
EPOC	Cochrane Effective Practice and Organisation of Care
ERIC	Education Resources Information Center
ETAT	Emergency Triage, Assessment and Treatment
GRADE	Grading of Recommendations Assessment, Development and Evaluation
ICC	Intra-class Correlation Coefficient
IMCI	Integrated Management of Childhood Illness
LICs	Low-Income Countries
LILACS	Latin American Caribbean Health Sciences
MoMS	Ministry of Medical Services
NLS	Newborn Life Support
NRP	Neonatal Resuscitation Program
PALS	Paediatric Advanced Life Support
RCT	Randomised Controlled Trial
SoF	Summary-of-Findings
WHO	World Health Organization
WHOLIS	World Health Organization Library Information System

List of papers

Paper I

Opiyo N, English M. [What clinical signs best identify severe illness in young infants aged 0-59 days in developing countries? A systematic review.](#) Arch Dis Child. 2011 Nov; 96(11):1052-9.

Paper II

Opiyo N, Were F, Govedi F, Gregan F, Wasunna A, English M. [Effect of newborn resuscitation training on health worker practices in Pumwani Hospital, Kenya.](#) PLoS One. 2008 Feb; 3(2):e1599.

Paper III

Opiyo N, English M. [In-service training for health professionals to improve care of the seriously ill newborn or child in low and middle-income countries \(Review\).](#) Cochrane Database Syst Rev. 2010, Issue 4. Art. No.: CD007071. DOI:10.1002/14651858.CD007071.pub2.

Paper IV

Opiyo N, Shepperd S, Musila N, Allen E, Nyamai R, Fretheim A, English M. Comparison of alternative evidence summary and presentation formats in clinical guideline development: a mixed-method evaluation. Submitted.

Paper V

Opiyo N, Shepperd S, Musila N, English M, Fretheim A. [The "Child Health Evidence Week" and GRADE grid may aid transparency in the deliberative process of guideline development.](#) J Clin Epidemiol. 2012 Sep;65(9):962-9.

Additional papers related to the dissertation

English M, Opiyo N. [Getting to grips with GRADE-perspective from a low-income setting](#). J Clin Epidemiol. 2011 Jul;64(7):708-10.

Opiyo N, English M. [Young infant sepsis: aetiology, antibiotic susceptibility and clinical signs](#). Trans R Soc Trop Med Hyg. 2007 Oct; 101(10):959-66.

Opiyo N, English M. [Newborn resuscitation: defining best practice for low-income settings](#). Trans R Soc Trop Med Hyg. 2006 Oct; 100(10):899-908.

Summary

Background

Despite recent overall improvement in the survival of under-five children worldwide, mortality among newborn babies remains high and accounts for an increasing proportion of child deaths in low-income countries. Decades of neonatal care research have resulted in identification of a range of simple low-cost interventions with potential to improve neonatal survival. However, many such evidence-based high-impact interventions are under-utilised in clinical practice.

Clinical practice guidelines are widely seen as important quality improvement tools that can contribute to delivery of evidence-based clinical practices. In this thesis we explore key processes aimed at improving neonatal care practices in rural hospitals in Kenya through the development and implementation of evidence-informed clinical practice guidelines.

Methods

We conducted five studies to address our objective. The first study, a systematic review, investigated a minimum set of diagnostic criteria (best clinical signs) for identifying severe young infant disease (a crucial first step in appropriate guideline-based management of severe disease). Severe illnesses were defined as cases warranting referral- or hospital-level care.

The process of implementing evidence-based recommendations into routine health worker practice was subsequently explored through two linked studies: a randomised controlled trial evaluating the effectiveness of an emergency neonatal resuscitation training course on health worker resuscitation practices in a hospital setting in Kenya; and a Cochrane systematic review evaluating the effectiveness of in-service emergency care training on health workers ability to manage seriously ill neonates or children in resource-poor countries.

The last two studies addressed different aspects of clinical practice guideline development:

Study four, a randomised controlled trial with a nested interview study, assessed the usefulness of alternative formats for summarising and presenting evidence for groups who are responsible for developing clinical practice guidelines. Healthcare professionals attending a one-week Kenyan, national guideline development workshop were randomly allocated to receive evidence packaged in three formats: (1) systematic reviews (SRs) alone; (2) systematic reviews with summary-of-findings tables (SRs with SoF tables); and (3) 'graded-entry' formats (a 'front-end' summary and a contextually framed narrative report plus the SR). The influence of format on ability to retrieve key outcome information, the primary outcome, was assessed using a written test. Interviews conducted within two months following completion of trial data collection explored panel members' views on the evidence summary formats and experiences with appraisal and use of research information.

The final study explored the translation process from evidence to care recommendations during a one week national guideline development workshop ('Child Health Evidence Week') in Kenya. Workshop discussions were aided by using a GRADE (Grading of Recommendations Assessment, Development and Evaluation) grid. Three investigators independently observed and recorded comments made during the workshop, focusing on discussions about research evidence, practitioner experiences and values, and context-specific issues that might influence acceptability and implementation of proposed care recommendations.

Results

Five prospective observational studies (N=17,506) were included in the systematic review of clinical signs of severe illnesses in young infants. Overall, moderate to high quality evidence indicated that, among sick infants aged 0 to 59 days brought to a health facility, the following clinical signs (alone or in combination) are likely to be the most valuable in identifying infants at risk of severe illness warranting hospital-level care: *history of feeding difficulty, history of convulsions, temperature (axillary) $\geq 37.5^{\circ}\text{C}$ or $< 35.5^{\circ}\text{C}$, change in level of activity, fast breathing/respiratory rate ≥ 60 breaths per minute, severe chest indrawing, grunting and cyanosis.*

In the randomised controlled trial of a training course (also included in the Cochrane review), data were collected on 97 and 115 resuscitation episodes in the intervention and control groups respectively. Trained providers demonstrated a higher proportion of adequate initial resuscitation steps compared to the control group (trained 66% versus control 27%; risk ratio 2.45, 95% confidence interval, CI 1.75 to 3.42). In addition, there was a statistically significant reduction in the frequency of inappropriate and potentially harmful practices per resuscitation in the trained group (trained 0.53 versus control 0.92; mean difference 0.40, 95% CI 0.13 to 0.66).

Two randomised controlled trials were included in the Cochrane review of effectiveness of in-service emergency care training courses. Overall, limited evidence from the included trials indicated that Newborn Resuscitation Training and Essential Newborn Care courses may result in short-term improvements in health worker resuscitation and delivery room neonatal care practices.

In the trial of different evidence summary formats, data were collected from sixty-five (93%) of 70 participants in the guideline development panel. There were no differences between the comparison groups in the odds of correct responses to key clinical questions (odds ratios, SRs with SoF tables versus SRs alone: 0.59, 95% CI 0.32 to 1.07; 'graded-entry' format versus SRs alone: 0.66, 95% CI 0.36 to 1.21). 'Graded-entry' formats were associated with a higher mean composite score for clarity and accessibility of information about the quality of evidence than SRs alone (mean difference 0.52, 95% CI 0.06 to 0.99). Findings from interviews with 16 of the 70 panelists indicated that short narrative evidence reports were preferred for the improved clarity of information presentation and ease of use.

Three key themes emerged in the final observational study of the guideline development process: (1) participants' 'referral to other evidence to support or refute the proposed care recommendations'; (2) participants' 'assessment of the presented research evidence'; and (3) participants' 'assessment of the local applicability of evidence'. Identified challenges to effective translation of evidence into recommendations included: absence of evidence, low quality or inconclusive evidence, inadequate reporting of key features of aspects of care under consideration, and differences in panelists' interpretation of the research literature.

Discussion

Early and accurate identification of severe young infant disease is a crucial step in appropriate management of severe illnesses in young infants. We identified an evidence-based diagnostic algorithm (comprising eight clinical signs) useful in the initial assessment and identification of severely ill infants warranting hospitalisation. To improve the diagnostic value of the algorithm, there is need for concurrent sensitisation of families on the eight signs (e.g. through community health education) to encourage early care seeking for probable serious illness. For one major condition, birth asphyxia, our findings indicate that a simple one day newborn resuscitation training course, adapted to local resources, may improve health workers capacity to provide adequate resuscitation at birth. However, the evidence for the effectiveness of emergency care courses on long-term health worker practices and clinical outcomes (e.g. neonatal survival) remain sparse – highlighting the need for further research on their true value in low-income settings.

Providing guideline development groups with relevant evidence that is appropriately packaged for their use is a particular challenge in the translation of evidence into contextually appropriate guideline recommendations. We addressed this knowledge gap in the fourth study, and demonstrated that 'graded-entry' evidence summary formats ('front-end' summary of key information linked to locally relevant factors that support implementation, as well as the full systematic review), may help those developing guidelines to access and contextualise research evidence. Finally, findings from our observations of panelists' discussions indicate that the process of the 'Child Health Evidence Week' combined with use of the GRADE grid instrument may improve transparency in the deliberative process of guideline development. With further refinement, this approach may provide an efficient and inclusive guideline development model for use in other low-income countries.

Conclusion

The findings presented in this thesis can have important implications for improving the quality and effectiveness of newborn care in low-income countries. Widespread implementation of the diagnostic algorithm and basic newborn resuscitation training courses may contribute to substantial reductions in neonatal mortality. The current work also provides important lessons on how research evidence can be packaged and presented to inform local care policies and practices. The identified challenges in the translation of evidence into guideline recommendations underscore the need to

improve skills in evidence-based medicine to support guideline development in low-income countries and create health system contexts that support recommended practices.

Introduction

In the arena of child health, neonatal care remains a relatively neglected area although it accounts for an increasing proportion of child deaths in low-income countries (LICs).¹⁻³ Research over the last decade on neonatal care has, however, resulted in identification of a range of apparently simple low-cost interventions of proven benefit and which if implemented at scale could lead to large reductions in neonatal mortality.⁴ An estimated 35 to 66% of neonatal deaths in LICs could be prevented if such interventions were implemented effectively with high coverage. Yet although such interventions are simple, surveys have revealed consistent under-utilisation of these high impact interventions and subsequently, a high prevalence of sub-standard neonatal care practices in LICs.⁵ Many of the barriers to provision of appropriate evidence-based neonatal care may be related to problems with accessibility and use of evidence by healthcare professionals. The realisation that failing to use research findings in health care has a negative impact on patient care, has led to an increased emphasis on finding and using appropriate ways of transferring research evidence into policy and practice.⁶⁻¹⁰

Clinical practice guidelines are widely seen as important quality improvement tools that can contribute to evidence-based clinical practices. In this thesis we explore key processes aimed at improving neonatal care practices in rural hospitals in Kenya through the development and implementation of evidence-informed clinical practice guidelines. To begin with, we systematically reviewed available evidence to support development and adaptation of Kenyan newborn and child health clinical practice guidelines. Specifically, the first study, a systematic review, investigated a set of minimum diagnostic criteria (best clinical signs) for severe young infant disease (a crucial first step in appropriate guideline-based management of severe disease).

The process of introducing evidence-based recommendations into routine emergency care was subsequently explored through two linked studies: a randomised controlled trial evaluating the effectiveness of an evidence-based emergency neonatal resuscitation training course on health worker resuscitation practices (Paper II); and a Cochrane systematic review evaluating the evidence base for short, targeted in-service evidence based emergency care training courses for the management of the severely ill newborn or child (Paper III).

The important intermediate step of translating evidence into contextually appropriate neonatal care recommendations was explored through two linked studies: a randomised controlled trial with a nested interview study evaluating the usefulness of alternative evidence summary and presentation formats for national guideline development groups (Paper IV); and an observational study exploring the translation process from evidence to care recommendations during a one week national guideline development workshop ('Child Health Evidence Week') (Paper V). A brief background and rationale for these studies is outlined below.

Clinical diagnosis of severe disease in young infants (Paper I)

Most young infant deaths in developing countries continue to occur in homes with unwillingness, inability or delay in care seeking precluding appropriate referral of severely ill infants to health facilities.¹¹ When healthcare is sought, primary and even secondary health facilities (rural hospitals) in resource-poor countries often have no specialists (such as paediatricians) and limited or no laboratory diagnostic capability.⁵ In such settings, clinical decisions for appropriate management of severely ill infants have to be made on the basis of presenting clinical signs and symptoms alone.

Which clinical symptoms and signs are the most useful in such settings for identifying serious illness in this vulnerable group of patients? The current Kenyan adaptation of the World Health Organization (WHO) Integrated Management of Childhood Illness (IMCI) algorithm recommends a panel of 15 clinical signs and symptoms for the identification of possible severe disease in young infants. Training health workers to identify large numbers of signs and then using an algorithm based on all these signs in busy clinics in resource-poor settings may not be feasible in practice. We therefore sought to summarise the evidence available on clinical predictors of serious illnesses to help define a likely minimum set of signs that would be most useful in revised Kenyan national guidelines for the hospital care component of IMCI, and potentially to broader child survival programs such as the WHO's IMCI approach. The set of diagnostic features proposed as a basic algorithm for initiating referral or empiric treatment should be feasible to implement as part of a revised IMCI strategy in first referral level facilities staffed by health workers with only basic training.

Implementing recommendations into routine health worker practice – the value of training in emergency care

Newborn resuscitation training (Paper II)

Birth asphyxia is estimated to cause 0.7 to 1.6 million deaths a year globally with 99% of these deaths occurring in low-income countries.¹¹ Overall deaths in the first 7 days of life account for 23% neonatal mortality, with prematurity adding to the burden attributable to asphyxia.¹¹ In Kenya the practical impact of such statistics is revealed in one study in a district hospital maternity department where death before discharge occurred in 1 of 33 babies born alive.¹² Prematurity and birth asphyxia were the main causes of these deaths.¹² Effective resuscitation could prevent some of these deaths as well as improve the outcomes of surviving asphyxiated babies.¹¹ However, provision of appropriate newborn resuscitation care is dependent on the presence of an adequately skilled health worker in the home or the facility. To date little attention has been paid to furnishing health workers with these skills and we have little idea what the best strategies are for improving current practices. We do however know that inappropriate, ineffective or dangerous forms of practice are widespread.^{5,13,14}

In higher income settings newborn resuscitation training courses have proliferated. Although these can be expensive little is known about the effect they actually have on health worker behaviour.¹⁵

Where studies on the effect of life support training for any age group have been done they focus mostly on knowledge and skill retention observed in simulated practice following course participation. Few studies have examined outcomes considered more useful such as morbidity, mortality or real-life clinical practices.¹⁶

In-service neonatal and paediatric emergency care courses (Paper III)

Training of health care providers is commonly viewed as an effective way of implementing clinical practice guidelines. In developing countries most deaths among seriously ill children who come into contact with referral level health services occur within 48 hours of being seen.¹⁷ It is possible that immediate good quality and effective care provided by health professionals could reduce the number of these deaths.¹⁸ Provision of appropriate care is, however, dependent on the presence of adequately skilled health personnel at the point of care delivery.¹¹ To improve health workers capacity to provide effective care for seriously ill newborns and children in low-income countries, a number of in-service training courses, mainly based on developed countries' models, are proposed.

These courses include: (1) neonatal life support courses (e.g. Newborn Life Support (NLS), Neonatal Resuscitation Program (NRP)); (2) paediatric life support courses (e.g. Paediatric Advanced Life Support (PALS), Paediatric Life Support (PLS)); (3) life support and emergency care elements within the Integrated Management of Pregnancy and Childbirth (e.g. Essential Newborn Care (ENC)); and (4) components of other in-service child health training courses that deal with care of serious illness (e.g. Emergency Triage, Assessment and Treatment (ETAT), Control of Diarrheal Diseases (CDDs) and Acute Respiratory Infections (ARIs)) (Table 1, Paper III). Although such formalised educational programs vary in origin, scope and target audience, they are typically aimed at in-service rather than pre-service training, and are short and intensive with a structured approach to the presentation of their clinical subject. These life support courses emphasize early recognition of neonatal and paediatric emergencies and prevention of cardio-respiratory arrest and mortality through resuscitation.

In-service training, however, costs both time and money. Apart from the high costs of providing such courses (recovered in high income countries often with high course fees), attendance at these courses often means that important staff (instructors and participants) are absent from their normal duties with potential disruption to patient care.¹⁵ In the hope that they might improve the quality of care in many low- and middle-income countries, considerable global efforts and investments have gone into their further development, refinement and adaptation to meet individual country needs.¹⁶ Yet despite these investments and the faith placed in them by many organizations and institutions, clear evidence of the effectiveness of these courses in improving health workers ability to manage seriously ill neonates or children appears lacking.

Exploring methods to support development of evidence-informed, contextually appropriate national guidelines for neonatal care

Summarising and presenting evidence (Paper IV)

The failure of uptake of agreed best practices at health worker level could be due to a failure in the mechanisms used to develop and implement clinical practice guidelines intended to provide recommendations on what comprises best practice. In particular, well conducted systematic reviews, using explicit methods to reduce bias, are key resources for translating the best research evidence into practice.^{19,20} However, despite their well established advantages,²¹ the time it takes to complete a systematic review can mean that they are not well timed to inform healthcare decisions. Furthermore, the use of technical language may deter non-research audiences from applying systematic review evidence to a local context.²¹

Thus, using evidence to inform healthcare decisions faces two challenges. First, research-based knowledge presented in the long format of a systematic review may not be accessed, understood and used. Second, systematic reviews of the best research evidence are not sufficient for sound decision-making in healthcare.²² Evidence on local contextual factors and values also needs to be taken into account. To support guideline development and other knowledge translation processes a number of approaches for tailoring and packaging scientific knowledge are currently in use (Box 1, Paper IV). These approaches aim to present research knowledge in clear and concise, reader-friendly formats, and may increase use of research knowledge in health care decision-making by relevant stakeholders. However, most, but not all, of these approaches have to date been initiated from high-income settings.

The evidence translation process (Paper V)

There is broad agreement that clinical practice guidelines should be ‘evidence-based’ but there has been less agreement on how to achieve this. In response to criticism that their process of guideline development has not always been made explicit,²³ the World Health Organization (WHO) recently indicated that wherever possible its guidance should be supported by rigorous reviews of the evidence including critical appraisal using the GRADE (Grading of Recommendations Assessment, Development and Evaluation) tool.²⁴ This approach (and others)²⁵⁻²⁸ recognize that the evidence must be viewed in the context of any relevant local evidence (e.g. microbial resistance-patterns), what is feasible in the local clinical setting and what is acceptable to intended users (healthcare professionals) and patients.

Integrating global research evidence with locally relevant evidence and contextual factors has, however, rarely been undertaken in an explicit or structured fashion.^{23,29-31} This is a particular problem in low- income countries. For newborn and child health, current guidance in low-income countries is mainly derived from that provided by WHO and its global partners (e.g. The United Nations Children’s Fund, UNICEF). The process of incorporating such guidance into national

guidelines, often referred to as 'adaptation', is rarely described, and the roles that value-based judgments and context-specific information play in developing recommendations are often not clear.

Aims and objectives

The overall aim of this project was to explore key processes aimed at improving neonatal care practices in rural hospitals in Kenya through the development and implementation of evidence-informed clinical practice guidelines.

The specific objectives were:

1. To employ a systematic review of evidence to support the development and adaptation of Kenyan newborn and child health clinical practice guidelines for the management of severely ill young infants by:
 - Conducting a systematic review of available evidence to identify a minimum set of signs and symptoms that should comprise a basic, minimum standard for knowledge, clinical assessment and management of severe disease in young infants (Paper I)
2. To explore the value of training in emergency care as a means to implementing best practices into routine health care by:
 - Conducting a randomised controlled trial evaluating the effectiveness of an evidence-based newborn resuscitation training course on health worker resuscitation practices (Paper II)
 - Conducting a systematic review of the literature assessing the effectiveness of short, in-service emergency-care training courses aiming to change health worker practices when dealing with the seriously ill newborn or child (Paper III)
3. To explore the process of translating evidence into contextually appropriate clinical practice guidelines for neonatal care in a low-income setting by:
 - Examining the usefulness of alternative approaches to summarising and presenting evidence that is provided to policy makers and health workers as the first stage of developing national clinical practice guidelines (Paper IV)
 - Exploring the process and perceived value of an approach ('Child Health Evidence Week') for developing national clinical practice guidelines (Paper V)

Methods

What clinical signs best identify severe illness in young infants aged 0 to 59 days in developing countries? A systematic review (Paper I)

Search strategy and selection criteria

Potential articles for inclusion were identified by direct searches of The Cochrane Library and MEDLINE (both from inception to November 2009). The searches were performed by combining MeSH (Medical Subject Headings) terms that are indicative of acute illnesses of interest (sepsis, bacteraemia), predictive of illness severity (signs, symptoms, clinical predictors, clinical markers) and indicative of target age group (neonates, infants and children). Further published and unpublished papers were sought by screening through bibliographies of identified articles and writing to authors of identified relevant papers. No language or time limits were applied in the search strategy.

Studies were included if they reported a set of clinical signs predictive of severe illnesses or mortality in young infants aged 0 to 59 days. Studies conducted in high-income countries (as defined by the World Bank³²) were excluded given the different spectrum and prevalence of severe illnesses. While a variety of definitions of 'severe illness' episodes have been suggested, in this review severe illnesses were defined as cases warranting referral- or hospital-level care. Both community- and outpatient-based prospective observational studies were considered. Two reviewers independently screened through the titles and abstracts of identified articles and applied the pre-defined selection criteria to assess their eligibility. Disagreements were resolved by discussion.

Assessment of quality of evidence

The strength of evidence (reflecting the appropriateness of the study design to answer the clinical question, the plausibility of prediction based on clinical signs, and the quality, quantity, and consistency of evidence) was independently assessed using the GRADE (Grading of Recommendations Assessment, Development and Evaluation) approach.²⁴ The approach classifies the quality of evidence (i.e. 'the extent to which one can be confident that an estimate of effect or association is correct') into four categories: high, moderate, low or very low (Table 1, Paper I). The unique features of GRADE include: (1) explicit, comprehensive criteria for downgrading and upgrading quality of evidence ratings; (2) explicit evaluation of the importance of outcomes; and (3) clear separation of quality of evidence from the strength of recommendations. The GRADE evidence profiles were prepared by one reviewer and verified independently by a second reviewer. Discrepancies in the quality ratings were resolved by discussion.

Effect of newborn resuscitation training on health worker practices in Pumwani Hospital, Kenya: a randomised controlled trial (Paper II)

Study setting

The study was conducted in Pumwani Maternity Hospital in Nairobi, Kenya. This is the main maternity facility for Nairobi, Kenya and provides delivery care to 17,000 women each year. The hospital has approximately 90 nurse/midwives (60 assigned to the labour ward and 30 to the theatre) primarily responsible for delivery care and newborn resuscitation with 14 on duty at any one time (8 labour ward, 6 theatre).

Participants and randomisation

We randomly assigned labour ward and theatre staff to either early or late training groups. Eligibility criteria for initial randomisation were: personal work plans for the 3 months post-randomisation that neither included leave of more than 2 weeks duration nor rotation to another work station; routine responsibility for newborn resuscitation; and provision of informed consent.

Intervention

The intervention was a one day Newborn Life Support (NLS) training. The form of training drew heavily on the one day UK Resuscitation Council training¹⁶ in form but was significantly adapted to the Kenyan setting where resources are limited. The one day course teaches an A (Airway), B (Breathing) and C (Circulation) approach to resuscitation laying down a clear step by step strategy for the first minutes of resuscitation at birth. It comprises focused lectures aimed at understanding the modern approach to resuscitation and practical scenario sessions using infant manikins to develop skills in airway opening, use of a bag-valve-mask device and chest (cardiac) compressions.

Outcome measures

The primary outcome for the study was the proportion of resuscitation episodes in which appropriate initial resuscitation steps were practiced as recommended in the Newborn Life Support training. The primary outcome was further classified into two levels: perfect (where the health worker entirely followed the training guideline) and adequate resuscitation with minor, clinically insignificant deviations from the training guideline. Secondary outcomes were: the frequency of inappropriate and/or potentially harmful practices which might confer a direct risk to the baby or an indirect risk through the delayed initiation of appropriate interventions.

Data collection

To capture data, trained observers worked a shift pattern to ensure at least one was present in the hospital continuously (spanning all 24 hours) until the estimated number of observations required by

our sample size calculations were achieved. The practice observation check list used was based on the resuscitation steps included in the training. Resuscitation observers were nursing students who had been specially trained as a group over 3 days to make structured observations on newborn resuscitation using role play and scenarios and a standardised checklist. Routine data on delivery outcomes, admissions to nursery and their causes and outcomes were collected retrospectively for the 6 months prior to the first training (June 2006), for a period of 3 months between the first training and training of the remaining staff (September 2006) and for 3 months after this.

Data analysis

Two investigators and Newborn Life Support training instructors, blinded to the health workers' identity or training status, independently assigned a score to each resuscitation episode (based on a review of all of the information on the observation sheet) using a 5 point scale, where 5 represented perfect resuscitation (Appendix S3, Paper II). We computed risk ratios (RRs) and 95% confidence intervals (CIs) (also adjusted for clustering) for the primary outcome. Confounding was explored for the categorical variables sex, years of experience and place of work (labour ward or theatre) by calculating stratified, cluster adjusted risk ratios.

In-service training for health professionals to improve care of the seriously ill newborn or child in low- and middle-income countries: a Cochrane review (Paper III)

Search strategy and selection criteria

We searched The Cochrane Register of Controlled Trials (CENTRAL) and several other sources for eligible published and unpublished articles. We also checked references of retrieved articles and reviews and contacted authors to identify additional studies. No date or language restrictions were applied in the searches. Randomised controlled trials, cluster-randomised trials, controlled clinical trials, controlled before and after studies and interrupted time series studies that evaluated the effects of in-service neonatal and paediatric emergency-care training on objectively measured professional practice, patient outcomes, health resource/services utilization or training costs in healthcare settings (not restricted to studies in low-income settings) were eligible for inclusion.

Assessment of quality of evidence

Two review authors independently assessed the risk of bias of included studies using the Cochrane Effective Practice and Organisation of Care (EPOC) group criteria for assessment of methodological quality of studies³³ and rated them into three classes: low (low risk of bias for all key domains), high (high risk of bias for one or more key domains) and unclear risk of bias (unclear risk of bias for one or more key domains). We assessed the overall quality of identified evidence using the GRADE (Grading of Recommendations Assessment, Development and Evaluation) system.²⁴ We resolved disagreements regarding the quality ratings through discussion.

Comparison of alternative evidence summary and presentation formats in clinical guideline development: a mixed-method evaluation (Paper IV)

Study design and participants

This was a mixed-method study incorporating a randomised controlled trial to assess the effectiveness of three different evidence summary formats with semi-structured follow-up interviews to explore panel members' views of these formats, experience with appraisal, use of and engagement with research evidence. Trial participants consisted of a multidisciplinary panel of healthcare professionals (N=70) who had been nominated to take part in a guideline development workshop. For the interviews, we purposively selected a sub-sample of participants (n=16).

Interventions

We assembled evidence in three formats: (1) systematic reviews (SR) alone (pack 'A'); (2) systematic reviews with summary-of-findings tables (SR with SoF tables; pack 'B'); and (3) 'graded-entry' formats (pack 'C'). Evidence pack 'A' represented the common standard practice of using systematic reviews and lengthy technical reports to inform healthcare policy and guideline development. Evidence pack 'B' represented the recently enriched format for preparing full Cochrane reviews.³⁴

The 'graded entry' format was designed to allow stepwise access to the evidence. It started with a 'front-end' short interpretation of the main findings and conclusions. These front-end concise summaries were linked to a locally prepared, short, contextually framed, 'narrative report'³⁵ in which the results of the systematic review were described and locally relevant factors that could influence the implementation of evidence-based guideline recommendations (e.g. resource capacity) were highlighted. The front-end summary and the narrative report were combined with the full systematic review (e.g. as published by the Cochrane Collaboration) to make a three-component set branded pack 'C'.

To prepare the summary-of-findings tables, we used the GRADE (Grading of Recommendations Assessment, Development and Evaluation) system²⁴ to appraise and summarise evidence. These tables were included in the front-end summaries, narrative reports and were available to support stand-alone systematic reviews. The summaries were delivered to participants as pre-reading materials one month before the workshop.

Randomisation

Evidence summaries in pack A, B and C formats were prepared for three 'tracer interventions': (1) feeding regimens in sick newborns³⁶; (2) hand hygiene for infection prevention³⁷; and (3) Kangaroo ('skin-to-skin') care for low birth weight babies.³⁸ We then provided all individual participants with evidence on all three tracer topics but used randomisation to ensure that all participants received one tracer-topic with packaging approach A, one with packaging approach B and one with packaging

approach C. In subsequent interviews we explored participants' views on all the 3 evidence packaging formats.

Outcome measures

The primary outcome was the proportion of correct responses to key clinical questions relevant to the specific tracer topics. These tested understanding of the effects of tracer-interventions on critical neonatal outcomes (mortality, morbidity). The secondary outcome measure was a composite score representing participants' self-reports of the clarity and accessibility of the evidence; participants rated their responses on a 3 to 5-point scale.

Data collection

Participants completed questionnaires on the first day (June 21st, 2010) of the guideline development workshop before the panel discussions about guidance recommendations. Participants were allowed up to 45 minutes to complete the questionnaire during which they had access to their personalized 'evidence packs'. Individual face-to-face interviews were conducted (between July and August 2010) by one investigator following completion of the questionnaire but before analysis. The interviews lasted approximately 30 to 45 minutes and focused on collecting information about participants' experiences with appraisal, the use of research evidence and views on evidence summary formats.

Analysis

The crude odds (likelihood) of correct responses for pack C compared to the odds for pack A (assumed baseline pack) were estimated using logistic regression. The secondary outcome measure was the 'clarity and accessibility' score. The mean 'clarity and accessibility' scores of pack C compared to the mean scores of pack A were estimated using linear regression. Similar processes were used to compare mean scores of pack B to A.

Audio interviews were transcribed verbatim by one investigator. Emerging themes and concepts were extracted by at least two co-investigators working independently.³⁹ These were compared and discussed; one investigator then summarised the recurrent concepts into a set of initial descriptive themes with narrative summaries explaining each theme. These were discussed by investigators iteratively, with reference to the original interview transcripts, until a final set of themes was agreed.

Exploring the evidence translation process during the development of evidence-based guidelines for newborn and paediatric care in Kenya: an observational study (Paper V)

Study design and participants

This was an observational study of the development of national guidelines for the management of common newborn and childhood illnesses in Kenya. Discussions among the participants at the national guideline development workshop ('Child Health Evidence Week') held between 21st and 25th, June 2010, were observed, recorded, analyzed and interpreted.

Child Health Evidence Week

During the 'Child Health Evidence Week' neonatal and paediatric stakeholders gathered to develop evidence-based guidelines for the hospital management of common newborn and childhood illnesses in Kenya.³⁵ Evidence summaries and draft recommendations were prepared in advance and sent to participants one month prior to the workshop. Participants were introduced to the GRADE system and the proposed procedures for the development of recommendations on the first day of the workshop. During the subsequent days the evidence underlying proposed recommendations was presented, using the PICO (Patient, Intervention, Comparator, Outcome) format and GRADE method to summarise the quality of the evidence and introduce possible additional considerations that might impact on the strength of recommendations. Each presentation was followed by a facilitated discussion. This initially focused on the formal evidence presented and subsequently, after presenting a draft recommendation, the wider issue of locally appropriate recommendations.

Draft recommendations were then amended where necessary and participants invited to vote for or against proposed recommendations. The voting process was aided by a modified GRADE grid,⁴⁰ a scaled polling table that allows participants to anonymously record their approval or disapproval of a proposed recommendation. Votes were counted and fed back to participants using PowerPoint to display bar graphs of the results allowing participants a final, short discussion prior to confirmation of a final recommendation. The presentation, discussion, revision of wording of recommendation and voting took approximately two hours for each of the clinical topics addressed. The deliberative process was facilitated by one, non-voting investigator.

Data collection

Three investigators independently observed and recorded comments made during the full workshop, focusing on discussions about research evidence, practitioner experiences and values, and context-specific issues that might influence acceptability and implementation of proposed recommendations.

Data analysis

Three investigators independently reviewed their field notes of panel discussions and grouped comments into a number of thematic clusters. Groupings were guided by previously identified criteria for assessing the applicability of systematic review evidence.⁴¹ Themes emerging from these initial, independent analyses were then compared and discussed iteratively among the investigators until a first common set of descriptive themes was identified. A table summarising these initial themes and sub-themes was then prepared and a final set of themes (supported by extracts from the field notes) was arrived at. Further exploratory analysis was completed by tabulating the frequency of aspects of the identified themes for each of the clinical topics.

Ethics

Papers I and III

Both were systematic reviews of existing research literature and required no ethical approval.

Paper II

Information about the implications, purpose and voluntary nature of participation was made available in written form to all labour ward and theatre staff and written informed consent was obtained from all health workers prior to their practice being observed. Information on the nature and purpose of the study and the need for the presence of an observer was also given to mothers admitted to the hospital for delivery. Mothers were given the opportunity to decline the presence of a resuscitation observer. Ethical approval for the conduct of the study was granted by the Kenya Medical Research Institute Scientific Committee and National Ethics Review Committee (Protocol No 1045).

Papers IV and V

Individual written informed consent for participation and audio recording of discussions was obtained prior to the face-to-face interviews. Confidentiality of participant information was ensured by assigning anonymous codes to individual audio interviews and transcripts. Ethical approval for the conduct of the studies was granted by the Kenya Medical Research Institute Scientific Committee and National Ethics Review Committee (Protocol No 1770).

Results

Paper I

Opiyo N, English M. [What clinical signs best identify severe illness in young infants aged 0-59 days in developing countries? A systematic review.](#) Arch Dis Child. 2011 Nov; 96(11):1052-9.

Aim: To summarise evidence from observational studies of clinical signs of severe illnesses in young infants aged 0 to 59 days, with a particular focus on defining a minimum set of best predictors of the need for hospital-level care.

Overall five prospective observational studies⁴²⁻⁴⁶ (N=17,506 infants) out of 404 identified papers were included in this review. All the included studies were conducted in resource-poor settings: three were based in outpatient clinics of first referral-level health facilities (basic or rural hospitals),^{42,44,45} one was community-based⁴³ while in another both outpatient and inpatient illness episodes were considered.⁴⁵ Three studies^{42,45,46} evaluated clinical predictors of severe illnesses while the remaining two^{43,44} reported risk factors for death.

Taken together, and based on the overlap of study results and the consistency of performance of clinical symptoms and signs in identifying severe illness, moderate to high quality evidence suggest that the following eight clinical signs (based on their strengths of associations (odds ratios) with severe illnesses, prevalence in the enrolled infants in the primary studies and ease of clinical recognition) are likely to be the most valuable in predicting severe illnesses in young infants presenting at primary healthcare facilities: *history of feeding difficulty, history of convulsions, temperature (axillary) $\geq 37.5^{\circ}\text{C}$ or $< 35.5^{\circ}\text{C}$, change in level of activity, fast breathing/respiratory rate ≥ 60 breaths per minute, severe chest indrawing, grunting and cyanosis* (Table 1, Paper I).

Paper II

Opiyo N, Were F, Govedi F, Gregan F, Wasunna A, English M. [Effect of newborn resuscitation training on health worker practices in Pumwani Hospital, Kenya.](#) PLoS One. 2008 Feb; 3(2):e1599.

Aim: To determine if a simple one day newborn resuscitation training alters health worker resuscitation practices in a public hospital setting in Kenya.

Two hundred and twelve resuscitation episodes were observed for 83 providers. We observed a significantly higher proportion of perfect initial resuscitation steps (24%) among trained providers compared to the control group (10%) (risk ratio, RR: 2.27, 95% CI 1.23 to 4.22; $p=0.009$). Similarly, the proportion of adequate initial resuscitation steps was higher among trained (66%) providers as compared to the control group (27%) (RR 2.45, 95% CI 1.75 to 3.42; $p<0.001$) (Table 2, Paper II). In addition, there was a statistically significant reduction in the frequency of inappropriate and potentially harmful practices per resuscitation in the trained group (trained 0.53 versus control 0.92;

mean difference 0.40, 95% CI 0.13 to 0.66). Group comparison for the overall mortality in all the resuscitation episodes showed no statistically significant differences between the groups (trained 0.28 (18/65), 95% CI 0.17 to 0.40; control 0.25 (9/25), 0.12 to 0.42; $p=0.77$).

Paper III

Opiyo N, English M. [In-service training for health professionals to improve care of the seriously ill newborn or child in low and middle-income countries \(Review\)](#). Cochrane Database Syst Rev. 2010, Issue 4. Art. No.: CD007071. DOI:10.1002/14651858.CD007071.pub2.

Aim: To investigate the effectiveness of in-service training of health professionals on their management and care of the seriously ill newborn or child in low- and middle-income settings.

In total, we identified 2,480 references from both the electronic and supplementary searches. We retrieved the full texts of 146 papers for further eligibility assessment. From these, we identified eight studies⁴⁷⁻⁵⁴ as potentially meeting the review inclusion criteria. We excluded six of these studies for a variety of reasons following a detailed assessment (e.g. due to inadequate study designs, non-enrolment of children with severe disease). Overall, we included two randomised controlled trials that met all the inclusion criteria.^{52,54} One of the two was our own trial reported on in Paper II.⁵² As a formal meta-analysis was not appropriate (given substantial differences in interventions and reported outcomes) a description of the results of the two trials is provided below.

In the first trial,⁵² newborn resuscitation training was associated with a significant improvement in performance of adequate initial resuscitation steps (risk ratio 2.45, 95% confidence interval, CI 1.75 to 3.42, $p<0.001$) and a reduction in the frequency of inappropriate and potentially harmful practices (mean difference 0.40, 95% CI 0.13 to 0.66, $p=0.004$). In the second trial,⁵⁴ available limited data suggested that there was improvement in assessment of breathing and newborn care practices in the delivery room following implementation of Essential Newborn Care (ENC) training course.

Paper IV

Opiyo N, Shepperd S, Musila N, Allen E, Nyamai R, Fretheim A, English M. Comparison of alternative evidence summary and presentation formats in clinical guideline development: a mixed-method evaluation. Submitted.

Aim: To assess the usefulness of different formats for summarising and presenting evidence for use in clinical guideline development.

Quantitative findings

Sixty-five (93%) of 70 panel members completed questions on primary outcome measures. There were no differences between the comparison groups in the odds of correct responses to key clinical

questions (odds ratios, systematic reviews with summary-of-findings tables (SRs with SoF tables) versus SRs alone: 0.59, 95% CI 0.32 to 1.07; 'graded-entry' format versus SRs alone: 0.66, 95% CI 0.36 to 1.21). 'Graded-entry' formats were associated with a higher mean composite score for clarity and accessibility of information about the quality of evidence than SRs alone (mean difference: 0.52, 95% CI 0.06 to 0.99).

Similarly, 'graded-entry' formats, compared to SRs alone, were associated with a 1.5 higher odds of judgments about the quality of evidence for critical neonatal outcomes being clear and accessible (adjusted OR: 1.52, 95% CI 1.06 to 2.20). There was no evidence that SRs with SoF tables improved this composite score compared to SRs alone (adjusted mean difference: 0.11, 95% CI to 0.71 to 0.48) (Table 3, Paper IV).

More than half of the respondents (60%) found systematic reviews to be more difficult to read compared to narrative reports, but some (17%) responded that systematic reviews were easy to read. About half of the participants (51%) found systematic reviews to be easier to read compared to summary-of-findings tables (26%). A higher proportion of participants preferred evidence summarised in narrative report formats to the full version of the systematic reviews (53% versus 25%) (Table 4, Paper IV).

Results from the interview study

Views on different formats for presenting systematic review evidence (see Panel 1, Paper IV, for illustrative quotes)

The majority of participants interviewed found research information summarised in the form of narrative reports to be clearer, easy to read, easy to understand, and containing 'just the right amount of information'. Conversely, participants expressed considerable variability in views for systematic reviews and summary-of-findings tables: while some found the comprehensive and structured nature of information presentation in systematic reviews to be useful, a number expressed difficulties with extracting pertinent information. Some participants found summary tables to be good for 'rapid consultation'; however, a number of participants found them difficult to understand as stand-alone summaries. Of note, many participants reported lack of time and the volume of evidence as factors contributing to inadequate engagement with the evidence.

Panelists' experiences with appraisal and use of research evidence (see Panel 2, Paper IV, for illustrative quotes)

The majority of participants responded that they were not conversant with assessing the quality of scientific literature or evidence-based medicine terminology, such as PICO (Patient, Intervention, Comparison, Outcome). A number suggested that a short course on evidence-based medicine would be beneficial to support evidence-based guideline development.

Paper V

Opiyo N, Shepperd S, Musila N, English M, Fretheim A. [The "Child Health Evidence Week" and GRADE grid may aid transparency in the deliberative process of guideline development.](#) J Clin Epidemiol. 2012 Sep;65(9):962-9.

Aim: To explore the evidence translation process during a one week national guideline development workshop ('Child Health Evidence Week') in Kenya.

Three key themes emerged from the field notes documenting participants' (N=70) discussions:

(1) Participants' 'referral to other evidence to support or refute the proposed recommendations'

Aspects of research evidence cited included: potential benefits and harms associated with treatment options, estimates of the magnitude of benefit associated with treatments, absence of relevant evidence and inconclusive evidence about the effectiveness of treatment options. Comments reflecting clinician experiences with proposed treatment options included experiences with routine clinical impacts of aspects of care, practical difficulties associated with treatments and patient acceptability of proposed treatments.

(2) Participants' 'assessment of the presented research evidence'

A range of issues reflecting participants' scrutiny of the credibility of the evidence was noted and included: sample size (power) issues, adequacy of available evidence on patient-relevant outcomes (e.g. mortality data), study execution (e.g. reliability of findings given premature study termination), adequacy of participant follow-up period, appropriateness of the study population (e.g. limited generalisability from recruiting inpatient populations) and measurement and selection of outcomes (e.g. potential biases associated with lack of blinding). Opinions diverged more frequently in clinical conditions where the quality of evidence was low.

Aspects of the nature of interventions discussed included: definitions of the interventions, intensity of interventions, descriptions of any co-interventions, techniques relating to how interventions were delivered and the content of interventions. Comments alluding to differences in participants' interpretation of evidence included various opinions regarding: sub-group of populations to which results apply, range of factors explaining differences in study results and outcome definitions.

(3) Participants' 'assessment of the local applicability of evidence'

Likely barriers and facilitators to effective implementation (adoption) of proposed treatment options cited included: costs of interventions, resource availability (including training), logistical issues, physical barriers, practical difficulties and compliance issues.

Comments where participants referred to knowledge about locally relevant practice-setting factors reflected awareness of local antimicrobial resistance patterns, local prevalence of febrile illnesses and nature of available clinical skills. Comments reflecting judgments about likely benefits and harms of alternative treatments included comparison of different types of benefits associated with treatments and considerations of benefits of treatments versus resource consumption. Lack of cost data for most interventions seemed to limit explicit judgments about their net value. References to health worker perspectives, attitudes, cultural issues, preferences and acceptability of proposed interventions were also observed.

Discussion

The five studies presented in this thesis addressed key processes aimed at improving neonatal care practices in Kenyan rural hospitals through the development and implementation of evidence-informed clinical practice guidelines. The first study, a systematic review, investigated and identified minimum diagnostic criteria for severe young infant disease (a crucial first step for appropriate guideline-based management of severe disease). In the second trial, we investigated the effectiveness of a simple one day evidence-based newborn resuscitation training course on health worker resuscitation practices. The third study, a Cochrane review, appraised available evidence on the effectiveness of short emergency care training courses on health worker management of severely sick newborns and children.

The final two studies addressed the methodology of clinical practice guideline development focusing on mechanisms for improving accessibility and use of evidence (Paper IV) and ensuring transparency, and minimising bias in the decision-making process of guideline development (Paper V).

The first part of this section focuses on how we attempted to minimise likely errors and biases in the conduct of the studies outlined above. In the second part I present a summary and interpretation of main findings from each of the five studies. I also compare our findings to related studies and reviews, highlight key study limitations, and discuss the implications of our findings to healthcare policy, and clinical practice. Finally, I present an overall integration of the individual study findings and discuss pertinent research gaps.

Internal validity

Causation and bias minimization

A key aim in most research studies is to assess whether there is a relationship between an exposure (e.g. targeted training) and an outcome (e.g. health worker practices). This process requires pre-specifying a hypothesis linking an exposure to an outcome (e.g. training and change in health worker practices). Causality is then established if we can demonstrate that observed results are unlikely to be due to 'chance', alongside other considerations, such as strength of the association, consistency of evidence from various research studies, presence of a dose-response relationship between the exposure and outcome, among others.⁵⁵ Traditionally, the randomised controlled design is used to reduce confounding and possible bias in the selection and assignment of participants to comparison study groups. In this thesis we successfully used both simple (Paper II) and stratified randomisation (Paper IV) to allocate participants to study interventions. This ensured fair comparisons between study groups (by enhancing balance of relevant prognostic factors) thus improving our confidence in the observed intervention effects. However, the 'classical' randomised controlled design is often not feasible or ethical; observational designs thus become necessary. In Paper I, a systematic review, we relied on evidence from observational studies of diagnosis and prognosis to define best clinical signs

of severe disease in young infants. We similarly adopted an observational design in Paper V to study the process of translating evidence into care recommendations during the 'Child Health Evidence Week'. Notably our structured 'real-time' observations of panelists' discussions provided additional insight on psychosocial factors that could influence decision making processes in multidisciplinary guideline groups (e.g. professional status) - thus enriching our interpretation of observed results, such as voting outcomes on draft recommendations.

Well conducted systematic reviews including meta-analyses are key tools for integrating existing evidence and establishing effectiveness and safety of therapeutic and health systems interventions. For example, systematic reviews may help establish whether the effects of interventions are consistent across populations and settings. Pooling results of individual trials in a meta-analysis improves power and precision of estimates of risks or treatment effects, and is particularly useful where small but significant effects or conditions with relatively low event rates are being investigated. We conducted two systematic reviews to generate reliable data needed for rational decision making about the clinical diagnosis of severe disease in young infants (Paper I) and the value of short emergency care training courses on health worker practices in low-resource settings (Paper III). However, like individual studies, results of systematic reviews may be vulnerable to a number of biases limiting the accuracy and reliability of reported findings. We therefore implemented a number of approaches to improve the validity of review findings. First, study selection, data extraction and quality assessment were all done in duplicate and in an independent manner minimising likely reviewer error and bias. Second, the use of the GRADE system added scientific rigor to the process of compiling and rating the quality of summarised evidence – by ensuring a comprehensive, transparent and integrated assessment and synthesis of key determinants of quality of evidence (e.g. risk of bias, consistency and precision of estimates of treatment effects, directness of evidence, likelihood of publication bias, among others). Finally, exhaustive literature searches (including grey literature sources) without any language restrictions reduced the risk of publication bias.

Sampling

Samples rather than whole populations are often used to assess effectiveness or safety of interventions (due to practical constraints, costs or ethical dilemmas associated with studying whole populations). Use of sample parameters to estimate population parameters (e.g. prevalence of a disease, measure of association between an exposure and disease) is however associated with some 'unavoidable' measurement error which may reduce our confidence in estimates of treatment effects. Thus, samples of adequate sizes ('power') are required to be able to detect differences in effects of interventions with reasonable precision, if indeed they exist.

In the randomised controlled trial (RCT) (Paper II) and mixed-method study (Paper IV, RCT component) we used random sampling techniques with *a priori* power calculations to minimise measurement (random) errors associated with use of sample estimates, and hence enhance internal validity (a pre-requisite of external validity). Conversely, purposive (non-random) sampling methods

were used in participant recruitment to ensure adequate representation of stakeholders involved in newborn and child health in Kenya (Paper IV, interview component), and to allow examination of behavioral influences in decision making during guideline development (Paper V).

Similarly, the results of systematic reviews may be affected by random errors (i.e. deviation from the true treatment effect due to sampling variation). Random errors result in imprecision in estimates of treatment effects (normally reflected in wide confidence intervals of estimates of effects). Results of small studies are more vulnerable to the influence of sampling variation and on average tend to be less precise. Typically, the effect of random errors is minimized by large sample sizes. In the context of systematic reviews, meta-analysis minimizes random errors by synthesizing results from as many trials as possible. We however did not undertake any meta-analysis in our systematic reviews, owing to: in Paper I, substantial differences in study settings (outpatient clinics, community settings, both outpatient and inpatient), diagnostic criteria for severe illnesses (expert clinician opinion, laboratory data) and outcome measures (clinical predictors of severe disease versus risk factors for death); and in Paper III, differences in interventions (newborn resuscitation course versus essential newborn care course) and outcome measures (initial resuscitation practices versus essential newborn practices at delivery).

External validity

This relates to ‘applicability’ or ‘generalisability’ of study findings to populations and settings beyond those studied.^{56,57} Studies included in this thesis were designed to generate evidence applicable to low-income settings such as Kenya. This is reflected in our choice of studied participants (patients), interventions and settings (both geographical and clinical) as outlined below.

Studies included in the two systematic reviews (Papers I and III) were all conducted in low- and middle-income countries (Paper I: Bangladesh, Bolivia, Ethiopia, Ghana, India, Kenya, Pakistan, Papua New Guinea, South Africa, The Gambia, The Philippines; Paper III: Kenya, Sri Lanka). In addition, the studies enrolled participants (patients) comparable to our target populations. For example, all the studies included in the systematic review of clinical recognition of severe illnesses enrolled representative samples of young infants, with prevalent disease conditions (sepsis, bacteraemia) in low-income countries.

In the newborn resuscitation trial (Paper II), we studied the effectiveness of a simple short (one day) newborn resuscitation training course whose implementation requires only basic equipment (i.e. ‘bag-valve-mask’ device rather than advanced intensive care facilities) and air (rather than oxygen) for newborn resuscitation. The studied ‘life saving’ resuscitation intervention is therefore suited for use in Kenya and similar settings where healthcare resources remain scarce.

Finally, we adopted a pragmatic approach (large panel size, N=70 participants; less deliberation time, 5 day meeting; locally prepared short evidence summaries) to guideline development (Papers IV and V). The approach was well suited for our large task (i.e. developing recommendations for

assessment, investigation and treatment of 11 major newborn and childhood conditions). In particular, the 'Child Health Evidence Week' integrated a variety of recommended practices for production of sound clinical guidelines and helped achieve a balance between our limited resources (technical skills) and urgent need for evidence-based guidance. We believe our processes could be usefully adapted to improve the rigor, relevance and uptake of future guidelines in other clinical areas and resource-poor contexts.

Discussion of main results

Clinical diagnosis of severe disease in young infants

Severe infections remain an important cause of morbidity and mortality among young infants in low-income settings. Prompt and accurate diagnosis of severe infections in young infants is a critical step in appropriate management of severely ill infants. Our systematic review identified eight symptoms and signs as best predictors of severe illnesses in infants aged 0 to 59 days in low-income settings: *history of feeding difficulty, history of convulsions, temperature (axillary) $\geq 37.5^{\circ}\text{C}$ or $< 35.5^{\circ}\text{C}$, change in level of activity, fast breathing/respiratory rate ≥ 60 breaths per minute, severe chest indrawing, grunting and cyanosis*. This diagnostic algorithm had a high sensitivity and reasonable specificity for identifying severe young infant disease, and is especially suited for use by first-line health workers working at first-level health facilities (district hospitals) in resource-poor countries.

The diagnostic value of the algorithm in confirming or excluding probable severe infection in young infants presenting to primary health care facilities could be improved through: (1) implementation of community-based educational interventions aimed at improving early care seeking for probable severe disease (e.g. danger-signs sensitisation health education for families)⁵⁸; and (2) in-service training of health workers on recognition of the clinical signs (e.g. to minimise variation in interpretation between health providers). Such measures should ideally be linked to efforts to improve empiric treatment, supportive care and access to healthcare providers with higher levels of training.

The main strength of our systematic review was the high quality of the included studies which ranged from moderate to high according to the GRADE system. There were three main limitations in the summarised evidence. First, a relatively small number of studies (N=5) were included. However the studies enrolled a large number of infants (N=17,506) which increases our confidence in the review findings. Second, the main aim of clinical algorithms is to identify severe illness so that appropriate treatment is initiated promptly; clinical predictors of death (reported in two studies^{43,44}) may therefore be of limited value - as they indicate advanced stages of disease during which treatment may be less likely to work. Finally, differences in the diagnostic criteria for severe illness across included studies may limit the reproducibility of the clinical features in practice, for example the diagnostic value of a clinical sign may vary depending on whether it's self-reported by parents or elicited through clinician questioning.

Three important research gaps were identified. First, we suggest further large observational validation studies to confirm the effectiveness of the proposed minimum set of eight clinical signs and symptoms. Second, we suggest that the approach's diagnostic performance among HIV-infected infants should be investigated.⁵⁹ Finally, improvements to this clinical approach, possibly by combining it with bio-markers of severe illness, should be examined.

Our findings have important implications for improving the management of serious infections in young infants presenting to primary health care facilities in resource-poor countries. The diagnostic algorithm comprises a restricted set of simple clinical signs which is likely to be easy and cheap to adopt in practice.

Newborn resuscitation training

Newborn resuscitation is an important component of interventions needed to reduce neonatal deaths in resource-poor countries. Unfortunately, most health workers in low-resource settings, where the burden of asphyxia-related deaths remain highest, lack the capacity to provide adequate newborn resuscitation. Findings from our second study indicate that one day newborn resuscitation training is associated with significant improvement in the performance of initial resuscitation steps, and reduction in the frequency of inappropriate and potentially harmful resuscitation practices. However, the training intervention had no effect on birth asphyxia related mortality, although the trial was not specifically powered with mortality as the primary outcome.

The resuscitation technique evaluated in the second study comprised appropriate positioning, drying and keeping the baby warm, assessing heart rate and respirations, and provision of assisted ventilation using a bag and mask device where needed. It should be possible to perform these resuscitation steps with only a minimum of equipment and without access to intensive care skills or facilities.^{60,61} Recent research findings have strengthened this assumption by demonstrating that suction in the presence of meconium and the use of oxygen are in most newborns unnecessary.^{60,62-65}

Retention of adequate resuscitation knowledge and skills among health providers and resuscitation instructors represents a major challenge in the provision of appropriate resuscitation, especially in settings where health providers infrequently resuscitate newborns. There are only few studies assessing the duration of effect of training. Worryingly these suggest a rapid and linear decay in cardio-pulmonary resuscitation skills, from as early as two weeks after training, with skills deteriorating to pre-training levels by one year.⁶⁶⁻⁷⁰ These findings indicate that programmes need not only provide initial training but consider sustainable systems for regular refresher training to safeguard against loss of resuscitation skills.

Our study had a number of limitations, however. First, we cannot exclude the possibility of cross-group contamination, although this would tend to reduce the apparent effect of the intervention. Second, we only observed practitioners for a short period after training and are unable to provide any information on the duration of the training effect. Future studies should assess the impact of neonatal resuscitation training on long-term health worker and clinical outcomes.

Our findings do however indicate that newborn resuscitation training, adapted to local resources, is one effective strategy for translating scientific knowledge on resuscitation into clinical practice.

Applied widely, particularly in rural settings, basic newborn resuscitation could result in substantial reduction in birth asphyxia related morbidity and mortality in low-resource settings.⁷¹

In-service neonatal and paediatric emergency care courses

Currently, a variety of emergency care courses are being promoted in low-income countries as a means to improving the quality of care provided to seriously ill newborns and children. We conducted a systematic review to determine whether such courses actually make a difference in health worker practices and patient morbidity and mortality in low-resource settings. Our systematic review identified two well-conducted studies on the impact of neonatal or paediatric in-service emergency care training. Limited evidence from the two studies suggests a beneficial effect in the performance of initial resuscitation practices in the short-term⁵² and delivery room newborn care practices.⁵⁴

The limited evidence available can be attributed to a number of factors: First, most studies were excluded because of poor study designs, for example, most were retrospective with likely biased results. Second, most existing studies have assessed the impact of emergency care courses on less direct outcomes, mainly health provider knowledge and skills, which may not reliably predict actual performance in clinical practice. Change in provider practice behavior, and consequently improvements in patient morbidity and mortality represent more reliable measures of training impact. Third, the lack of well conducted studies could also be attributed to methodological and ethical challenges inherent in the evaluation of the impact of emergency care courses on clinician practice behavior and patient outcomes. For example, protection against 'contamination' of intervention effects to the control group can be difficult to avoid within routine practice settings. Furthermore, random assignment of health providers and seriously ill infants to a control arm and observation of practices performed by untrained providers clearly raises ethical concerns.

The main strength of our review was the exhaustive literature searches for relevant studies. A limitation of our review was the few number of methodologically strong studies included. Future studies in this area should use appropriate controls, and adequate randomisation procedures, and focus on long-term provider practices and patient outcomes as measures of training impact. Such studies should also collect data on resources used and costs of training interventions to allow assessments of whether they are cost-effective.

The findings of our review, in common with previous related reviews,^{15,72} demonstrate the sparse evidence base for the effectiveness of emergency neonatal and paediatric courses in low-resource settings. Despite this, these courses continue to be popular within ministries of health and healthcare institutions and are increasingly being promoted by influential groups such as the World Health Organization. Before these become the standard of care, making them even more difficult to evaluate, evidence of their ability at least to change health worker practices and ideally to reduce mortality are needed.

Evidence summary formats in guideline development

Best formats for summarising and presenting evidence for use in clinical guideline development remain unclear. Findings from the randomised trial component of our mixed-method study showed that 'graded-entry' evidence summary formats (a 'front-end' summary of key information and a contextually framed narrative report plus the full systematic review) may improve clarity and accessibility of research evidence in clinical guideline development. However, no effect on correct understanding of key evidence summary messages, our primary outcome, was found. The lack of what was hoped would be an improved format on ability to locate (and arguably comprehend) key messages on effect directions and size leaves the question of how to achieve this outcome open. It is possible that the lack of effect was due to inadequate study power to detect a small but useful effect. Alternatively inadequate participant engagement with the evidence summaries prior to the workshop may have been the problem - most participants reported spending minimal time reviewing the evidence summaries prior to the workshop.

The mixed-findings on the impact of our evidence summary formats and findings from the interview data showing difficulties accessing and interpreting evidence more generally perhaps highlight the need to strengthen guideline panelists skills in evidence-based medicine as a precursor improving translation of evidence into policy.

Importantly, we feel the 'graded-entry' summary formats and their components were reported to improve clarity of information presentation, improve accessibility to key information, be more reader-friendly (both narrative reports and summary-of-findings tables) and were preferred by participants over full systematic reviews provided alone. These favorable findings on secondary outcomes were further supported by the interview data. A preference for narrative reports may be due to: their abbreviated and plain language nature; incorporation of judgments on the quality of evidence for guideline relevant outcomes; and inclusion of contextual information (e.g. local antimicrobial resistance patterns). However, clearly some challenges remain in presenting evidence to those charged with developing policy in contexts such as Kenya. Future strategies worthy of consideration might include, for example, greater involvement of panel members in the preparatory stages of guideline development (e.g. in clinical question formulation and conduct of systematic reviews),^{29,73} or training of panel members on research methodology and guideline development methods (e.g. on the GRADE system). Additionally, to ensure guideline panelists effectively review pre-workshop materials (research summaries), sending e-mail or text message reminders, or introducing pre-meeting knowledge assessments might be considered.

In summary, although our findings suggest possible benefits of graded-entry formats, the best formats for improving actual understanding of research evidence and improving its use in guideline development processes appear to be lacking. Further research in this area should focus on the relative effectiveness of the various evidence summary formats that are available and how these may be combined with alternative approaches (e.g. GRADE evidence profiles) to effectively convey research evidence and contextual factors to guideline panelists.

Translation of evidence into recommendations

A common challenge in guideline development is the lack of transparency in the decision making process. We undertook structured observations of panelists' discussions during a one week guideline development workshop (the 'Child Health Evidence Week') in Kenya. An anonymous voting tool, GRADE grid instrument (Appendix C, Paper V), was used by the panel to reach consensus on the draft recommendations. Our findings give insight into the many factors that influence the decision making process in multidisciplinary guideline development, including research evidence, implementation factors, and practitioners' experiences and values.

Our observations also showed that participants actively engaged with the research evidence that was presented, and critically assessed its applicability, even though a large number of topics were covered in a relatively short period. It is however known that even when presented with an identical evidence base, different guideline panels may propose contrasting recommendations for the same clinical condition. This perhaps reflects, among other considerations, differences in panelists' interpretation of research literature.⁷⁴ Such differences may also be a barrier to timely achievement of consensus in multidisciplinary panels. Furthermore, substantial differences in interpretation of evidence may weaken the validity of care recommendations, and compromise their acceptability by intended users. No proven methods currently exist on how best to overcome such problems. Given these caveats, the process adopted (Child Health Evidence Week) resulted in the relatively rapid production and revision of multiple national recommendations in a transparent process that took account of recent research evidence and locally relevant data.

Integration of findings and implications for future research

On the whole, the research presented in this thesis spans key areas of evidence synthesis to support clinical guideline development, effects of specific health worker training linked to implementation of emergency care guidelines, and the translation process linking evidence synthesis to guideline recommendations. Taken together, our findings highlight a number of challenges that hinder effective development of, and delivery of evidence-based policies needed to improve the quality of newborn and child health in Kenya, and similar low-resource settings: First, if research evidence is to contribute to reductions in newborn and child mortality it must impact on the continuum of care, including, improved recognition of severe illness and high coverage of appropriate management at all levels of the health system. In low-income settings such as Kenya there remain challenges in all of the steps on this continuum. Few systematic reviews are conducted by researchers in low-income settings, the relevant research base to review is often limited and may be of low quality, and until recently there has been little attention to methods for reviewing important areas of clinical care such as optimizing diagnoses of common infections and treatment implementation.²⁹ There is therefore a continued need to improve capacity in low-income settings for conducting systematic reviews to support evidence-based practice and a need for mechanisms that engage low-income countries in identifying priority areas for review, weighing the evidence to formulate guideline recommendations, and, where needed, identifying appropriate primary research.⁷⁵ Such efforts to strengthen capacity for research synthesis, and cooperation in the production and use of research evidence in guideline development processes represent a first important step in bridging the gap between evidence and its implementation in clinical practice.

Second, research synthesis is only an initial step in the process of improving patient care. Such evidence should then be used to inform development of contextually appropriate guideline recommendations for practitioners. In the past most clinical recommendations in low-income countries were simple replicas of World Health Organization guidance (which itself may be flawed²³) reproduced in countries after an opaque process of ‘adaptation’ – usually an unstructured meeting of a few subject experts. For example, in the case of the WHO ‘Pocketbook’ (www.ichrc.org), and specifically its advice on neonatal care, we are not aware of any reports of how the influence of local health systems capacity and economic factors have been taken into account in producing national guidelines for hospital care.

While many developed nations, for example the UK’s National Institute for Health and Clinical Excellence (NICE), have invested considerable sums in formalizing the development of clinical guidelines and promoting wide participation, including from health system users or patient groups, such approaches are likely to be too costly for most low-income countries for many years to come. We therefore developed and examined a unique, relatively low-cost way of developing locally acceptable and relevant consensus care recommendations with limited resources. This differed in a number of ways to guideline development structures used in established institutions such as NICE and others.^{27,76} We deliberately adopted, unlike these higher income institutions, a larger panel (70 participants), broader scope (11 clinical topics), shorter deliberative time frame (5 days meeting),

and GRADE grid to synthesize individual views and reach consensus. This approach recognized the limited resources and technical capacity available within Kenya balanced against the need for developing a potentially sustainable mechanism for making (and updating) evidence-informed guidelines within Kenya. While there were clear weaknesses of this approach, for example the absence of patients' views in the process, the strategy fostered much wider engagement in developing national care recommendations than any previous exercise. Arguably this resulted in a greater sense of ownership of the final recommendations, a greater appetite for evidence and a greater appreciation of the value of evidence for decision making. The strategy is, we argue, an improvement on prior approaches to national guideline development. Traditionally, countries such as Kenya have simply relied on generic guidelines provided by organizations such as the World Health Organisation that are adapted, in a rather opaque 'expert' processes, for use at national level. Encouraged by this initial experience we therefore suggest the approach of the 'Child Health Evidence Week' combined with the GRADE grid may be further explored and developed to provide an efficient and inclusive rapid guideline development model for use in other low-income countries.

Beyond developing guidance there are a number of challenges in translating recommendations into actual practice by clinicians at the point of care. Part of the problem may be that most clinical practice guidelines continue to be assembled as lengthy books / booklets which limits their effective use in practice. In Kenya, brief, 32 page guideline booklets used as part of a multifaceted package (including training, supervision, local facilitation and face-to-face feedback) helped improve uptake of guideline recommended practices in Kenyan hospitals for multiple childhood diseases.⁷⁷ However, although the guideline booklets also included neonatal guidelines, evidence of a 'spill-over' effect of the paediatric care intervention to improve neonatal care practices was limited.⁷⁸ More research would therefore be useful to identify guideline formats that facilitate adoption of recommendations during typical patient encounters (e.g. 'user-friendly' electronic presentation formats) and on how these may be combined with further reinforcement interventions in low-income settings to improve newborn care.

Training of health workers may be an important component of strategies for promoting effective implementation of guidelines in clinical practice. Although there is often undue reliance on this intervention alone when it may have limited value.⁷⁹ In this thesis we have demonstrated that health workers do need improved knowledge and skills in the area of neonatal resuscitation and the potential of a basic resuscitation training course to improve health worker adherence to recommended newborn resuscitation practices at birth. However, we also note that this single training course did not result in perfect uptake of recommendations while other literature suggests training effects can be short-lived. Thus mechanisms are needed for ongoing reinforcement of good practices, a subject rarely tackled in low-income settings. Possible strategies include implementing effective routine quality assessment tools that provide measurement of the uptake and impact of guideline implementation linked to re-training or other reinforcement activities. Another approach that may be useful is introduction of criterion-based clinical audits (CBCAs). These involve evaluating delivery of 'best practices' by extracting relevant data from patient records and comparing the care

provided against agreed guideline-based standards of care.⁸⁰ CBCAs have been shown to be feasible and effective tools for bridging the quality gap in some developing countries.⁸¹

For emergency care of severely ill infants, the research base for the effects of training courses in resource-poor countries remains weak. The weak research base is due in large part to the methodological challenges in evaluation of educational interventions targeting severely ill children in routine practice settings with traditional randomized controlled trial designs (e.g. ethical dilemmas associated with random allocation of patients to health providers not trained in emergency care and high likelihood of cross-group contamination making it difficult to establish training effectiveness where this truly exists).⁸² For example in our newborn resuscitation trial (Paper II) our criteria for randomisation, aiming to ensure health workers were present to be observed in a defined period of three months, resulted in few staff being eligible (raising the possibility of bias in group allocation).

A further challenge to effective application of guideline recommendations in practice settings is the comorbid nature of most newborn and childhood illnesses. Guidelines typically address single conditions (e.g. malaria) yet most children in practice present with multiple illness episodes (e.g. malaria and bacteraemia).^{83,84} Comorbid conditions may limit effectiveness of treatments for index conditions. Further research is needed on how to develop treatment recommendations that take account of comorbid conditions.

Once developed it is clear that guideline recommendations need to be explicitly communicated and actively disseminated (e.g. using opinion leaders combined with frequent performance audit and feedback⁸⁵) to enhance their intended impact on practitioner behaviors. Clinical reminders⁷⁹ and educational outreach^{86,87} have been demonstrated to improve uptake of guideline recommendations. However, evidence on the effectiveness of these strategies has to date mainly come from high-income countries. More research is needed to validate these findings in low-income settings. Importantly, efforts to improve effective development and implementation of clinical guidelines need to be matched with concurrent health systems strengthening to maximize their impact on patient care and outcomes.

Clearly, no single strategy is likely to be successful in ensuring effective implementation of scientific evidence. Based on the findings of studies presented in this thesis, we propose an integrated approach to improve and sustain the uptake and use of findings from neonatal and child health research, encompassing: increasing practitioners' participation in research studies and guideline development processes, appropriate packaging of evidence in accessible and easy-to-use formats, specific training interventions addressing policymakers' and clinicians' 'skill-gaps' in evidence-based practices and inclusion of updated evidence about effective interventions in pre-service and in-service education programmes. These should be combined with health system interventions that promote adoption of recommended health worker behaviours at scale. Finally, while these strategies address the supply side of health care measures to improve newborn and child survival in low-income countries need to also attend to social and cultural practices that may be harmful and promote care seeking on the 'demand side'.

Conclusion

The findings presented in this thesis can have important implications for improving the quality and effectiveness of newborn care in low-income countries. Widespread implementation of the diagnostic algorithm and basic newborn resuscitation training courses may contribute to substantial reductions in neonatal mortality. The current work also provides important lessons on how research evidence can be packaged and presented to inform local care policies and practices. The identified challenges in the translation of evidence into guideline recommendations underscore the need to improve skills in evidence-based medicine to support guideline development in low-income countries and create health system contexts that support recommended practices.

References

1. Central Bureau of Statistics & Ministry of Planning and National Development (2008). Kenya demographic and health survey - preliminary report. Nairobi: Government of Kenya.
2. Joy L, Kate K. Opportunities for Africa's newborns. Practical data, policy and programmatic support for newborn care in Africa. Geneva, Switzerland: The Partnership for Maternal, Newborn and Child Health. 2006.
3. Lawn J, Cousens S, Zupan J; Lancet Neonatal Survival Steering Team. 4 million neonatal deaths: When? Where? Why? *Lancet* 2005;365:891-900.
4. Darmstadt G, Bhutta Z, Cousens S, Adam T, Walker N, de Bernis L; Lancet Neonatal Survival Steering Team. Evidence-based cost-effective interventions: how many newborn babies can we save? *Lancet* 2005;365:977-988.
5. English M, Esamai F, Wasunna A, Were F, Ogutu B, Wamae A, et al. Delivery of paediatric care at the first-referral level in Kenya. *Lancet* 2004;364:1622-9.
6. Graham I, Logan J, Harrison M, Straus S, Tetroe J, Caswell W, et al. Lost in knowledge translation: time for a map? *J Contin Educ Health Prof* 2006;26:13-24.
7. Innvaer S, Vist G, Trommald M, Oxman A. Health policy-makers' perceptions of their use of evidence: a systematic review. *J Health Serv Res Policy* 2002;7:239-44.
8. Lavis J, Lomas J, Hamid M, Sewankambo N. Assessing country-level efforts to link research to action. *Bull World Health Organ* 2006;84:620-8.
9. Nutley S, Walter I, Davis H. From knowing to doing: a framework for understanding the evidence-into-practice agenda. *Evaluation* 2003;9:125-48.
10. Ward V, House A, Hamer S. Developing a framework for transferring knowledge into action: a thematic analysis of the literature. *J Health Serv Res Policy* 2009;14:156-64.
11. World Health Organisation. The World Health Report 2005: Make every mother and child count. Geneva, Switzerland: WHO, 2005.
12. English M, Muhoho A, Aluda M, Were S, Ross A, Peshu N. Outcome of delivery and cause-specific mortality and severe morbidity in early infancy: a Kenyan district hospital birth cohort. *Am J Trop Med Hyg* 2003;69(2):228-232.
13. Kamenir S. Neonatal resuscitation and newborn outcomes in rural Kenya. *J Trop Pediatr* 1997;43:170-173.
14. World Health Organization. The World Health Report 1995. Bridging the gap. 1995.
15. Jabbour M, Osmond M, Klassen T. Life support courses: are they effective? *Ann Emerg Med* 1996;28:690-698.
16. Baskett P, Nolan JP, Handley A, Soar J, Richmond S; European Resuscitation Council. European resuscitation council guidelines for resuscitation 2005. Section 9. Principles of training in resuscitation. *Resuscitation* 2005;67S1:S181-9.
17. Berkley J, Maitland K, Mwangi I, Ngetsa C, Mwarumba S, Lowe BS, et al. Use of clinical syndromes to target antibiotic prescribing in seriously ill children in malaria endemic area: observational study. *BMJ* 2005;330(7498):995.
18. Nolan T, Angus P, Cunha AJ, Muhe L, Qazi S, Simoes EA, et al. Quality of hospital care for seriously ill children in less-developed countries. *Lancet* 2001;357(9250):106-10.
19. Mulrow C. Rationale for systematic reviews. *BMJ* 1994;309(6954).
20. Tugwell P, Robinson V, Grimshaw J, Santesso N. Systematic reviews and knowledge translation. *Bull World Health Organ* 2006;84(8):643-51.

21. Lavis J, Davies H, Oxman A, Denis J, Golden-Biddle K, Ferlie E. Towards systematic reviews that inform health care management and policy-making. *J Health Serv Res Policy* 2005;10 (Suppl 1):35-48.
22. Coulter I. Evidence summaries and synthesis: necessary but insufficient approach for determining clinical practice of integrated medicine? *Integr Cancer Ther* 2006;5:282.
23. Oxman A, Lavis J, Fretheim A. Use of evidence in WHO recommendations. *Lancet* 2007; 369: 1883-89.
24. Guyatt G, Oxman AD, Vist G KR, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ, et al. Rating quality of evidence and strength of recommendations GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924-926.
25. National Institute for Health and Clinical Excellence (January 2009). The guidelines manual. London: National Institute for Health and Clinical Excellence. Available from: www.nice.org.uk
26. SIGN 50: a Guideline developer's handbook. Guideline Number 50. Revised Edition. <http://www.sign.ac.uk/pdf/sign50.pdf> (Accessed August 31, 2011).
27. Brouwers M, Kho M, Browman G, Burgers J, Cluzeau F, Feder G, et al. AGREE Next Steps Consortium. AGREE II: advancing guideline development, reporting and evaluation in health care. *CMAJ* 2010;182(18):E839-42.
28. Howick J. Levels of Evidence. Centre for Evidence-Based Medicine (CEBM). Oxford: University of Oxford, 2010. <http://www.cebm.net/index.aspx?o=5653> (Accessed August 31, 2011).
29. English M, Opiyo N. Getting to grips with GRADE-perspective from a low-income setting. *J Clin Epidemiol* 2011;64(7):708-710.
30. Raine R, Sanderson C, Black N. Developing clinical guidelines: a challenge to current methods. *BMJ* 2005;17:631-633.
31. Scott I, Guyatt G. Clinical practice guidelines: the need for greater transparency in formulating recommendations. *Med J Aust* 2011;195(1):29-33.
32. World Bank 2006. Data and Statistics, World Bank. <http://web.worldbank.org/> (Accessed 4 December 2009)
33. EPOC 2009. Resources for review authors: risk of bias criteria. Available at: <http://epoc.cochrane.org/sites/epoc.cochrane.org/files/uploads/Risk%20of%20Bias%2005-01-2009.doc> (Accessed 4 May, 2012).
34. Rosenbaum S, Glenton C, Oxman A. Summary-of-findings tables in Cochrane reviews improved understanding and rapid retrieval of key information. *J Clin Epidemiol* 2010;63:620-626.
35. Improving delivery of hospital care in Africa (iDOC Africa). Available at: <http://www.idoc-africa.org/etat-evidence-base/child-health-evidence-week-chew-2010.html> (Accessed April 4, 2012).
36. Edmond K, Bahl R. Optimal feeding of low-birth-weight infants. Technical review. World Health Organisation, 2006.
37. World Health Organization (WHO). WHO guidelines on hand hygiene in health care. Geneva: World Health Organisation, 2009.
38. Conde-Agudelo A, Belizán J. Kangaroo mother care to reduce morbidity and mortality in low birthweight infants. *Cochrane Database Syst Rev* 2003, Issue 2. Art. No.: CD002771. DOI: 10.1002/14651858.CD002771 2003.
39. Pope C, Ziebland S, Mays N. Qualitative research in health care. Analysing qualitative data. *BMJ* 2000;320(7227):114-116.

40. Jaeschke R, Guyatt H, Dellinger P, Schunemann H, Levy M, Kunz R, et al. Use of GRADE grid to reach decisions on clinical practice guidelines when consensus is elusive. *BMJ* 2008;337:a744.
41. Lewin S, Lavis J, Oxman A, Bastías G, Chopra M, Ciapponi A, et al. Supporting the delivery of cost-effective interventions in primary health-care systems in low-income and middle-income countries: an overview of systematic reviews. *Lancet* 2008;372(9642):928-939.
42. Young Infants Clinical Signs Study Group. Clinical signs that predict severe illness in children under age 2 months: a multicentre study. *Lancet* 2008;371:97-8.
43. Bang A, Bang R, Reddy M, Baitule S, Deshmukh M, Paul V, et al. Simple clinical criteria to identify sepsis or pneumonia in neonates in the community needing treatment or referral. *Pediatr Infect Dis J* 2005;24:335-41.
44. Duke T, Oa O, Mokela D, Oswyn G, Hwaihwanje I, Hawap J. The management of sick young infants at primary health centres in a rural developing country. *Arch Dis Child* 2005;90:200-5.
45. English M, Ngama M, Mwalekwa L, Peshu N. Signs of illness in Kenyan infants aged less than 60 days. *Bull World Health Organ* 2004;82:323-9.
46. Weber M, Carlin J, Gatchalian S, Lehmann D, Muhe L, Mulholland E, WHO Young Infants Study Group. Predictors of neonatal sepsis in developing countries. *Pediatr Infect Dis J* 2003;22:711-16.
47. Bryce J, Gouws E, Adam T, Black R, Schellenberg J, Manzi F, et al. Improving quality and efficiency of facility-based child health care through Integrated Management of Childhood Illness in Tanzania. *Health Policy Plan* 2005;20 (Suppl 1):i69-i76.
48. El Arifeen S, Blum L, Hoque D, Chowdhury E, Khan R, Black R, et al. Integrated Management of Childhood Illness (IMCI) in Bangladesh: early findings from a cluster-randomised study. *Lancet* 2004;364(9445):1595-602.
49. González Ochoa E, Armas Pérez L, Bravo González JR, Cabrales Escobar J, Rosales Corrales R, Abreu Suárez G. Prescription of antibiotics for mild acute respiratory infections in children. *Bull Pan Am Health Organ* 1996;30(2):106-17.
50. Gouws E, Bryce J, Habicht J, Amaral J, Pariyo G, Schellenberg J, et al. Improving antimicrobial use among health workers in first level facilities: results from the multi-country evaluation of the Integrated Management of Childhood Illness strategy. *Bull World Health Organ* 2004;82(7):509-15.
51. Nadel F, Lavelle J, Fein J, Giardino A, Decker J, Durbin D. Teaching resuscitation to pediatric residents: the effects of an intervention. *Arch Pediatr Adolesc Med* 2000;154(10):1049-54.
52. Opiyo N, Were F, Govedi F, Fegan G, Wasunna A, English M. Effect of newborn resuscitation training on health worker practices in Pumwani Hospital, Kenya. *PLoS One* 2008;3(2):e1599.
53. Pelto G, Santos I, Gonçalves H, Victora C, Martines J, Habicht J. Nutrition counseling training changes physician behavior and improves caregiver knowledge acquisition. *J Nutr* 2004;134(2):357-62.
54. Senarath U, Fernando D, Rodrigo I. Effect of training for care providers on practice of essential newborn care in hospitals in Sri Lanka. *Obstet Gynecol Neonatal Nurs* 2007;36(6):531-41.
55. Schünemann H, Hill S, Guyatt G, Akl E, Ahmed F. The GRADE approach and Bradford Hill's criteria for causation. *J Epidemiol Community Health* 2011;65(5):392-5.
56. Atkins D, Chang S, Gartlehner G, Buckley D, Whitlock E, Berliner E, et al. Assessing applicability when comparing medical interventions: AHRQ and the Effective Health Care Program. *J Clin Epidemiol* 2011;64(11):1198-207.
57. Rothwell P. External validity of randomised controlled trials: "to whom do the results of this trial apply?" *Lancet* 2005;365(9453):82-93.

58. Bazzano A, Kirkwood B, Tawiah-Agyemang C, Owusu-Agyei S, Adongo P. Beyond symptom recognition: care-seeking for ill newborns in rural Ghana. *Trop Med Int Health* 2008;13:123-8.
59. Carrol E, Mankhambo L, Jeffers G, Parker D, Guiver M, Newland P, et al. The diagnostic and prognostic accuracy of five markers of serious bacterial infection in Malawian children with signs of severe infection. *PLoS One* 2009;4:e6621.
60. World Health Organization. Basic newborn resuscitation: a practical guide. Geneva: World Health Organisation. 1997.
61. International Liaison Committee on Resuscitation. 2005 International consensus on cardiopulmonary resuscitation and emergency cardiovascular care science with treatment recommendations. Part 7: Neonatal resuscitation. *Resuscitation* 2005; 67.
62. Halliday H, Sweet D. Endotracheal intubation at birth for preventing morbidity and mortality in vigorous, meconium-stained infants born at term. *Cochrane Database Syst Rev* 2001, Issue 1. Art. No.: CD000500. DOI:10.1002/14651858.CD000500. 2001.
63. Opiyo N, English M. Newborn resuscitation: defining best practice for low-income settings. *Trans R Soc Trop Med Hyg* 2006;100:899-908.
64. Tan A, Schulze AA, O'Donnell CPF, Davis PG. Air versus oxygen for resuscitation of infants at birth. *Cochrane Database Syst Rev* 2004, Issue 2. Art. No.: CD002273. DOI: 10.1002/14651858. CD002273.pub3.
65. Vain N, Szyld E, Prudent L, Wiswell T, Aguilar A, Vivas N. Oropharyngeal and nasopharyngeal suctioning of meconium-stained neonates before delivery of their shoulders: multicentre, randomised controlled trial. *Lancet* 2004;364:597-602.
66. Berden H, Willems F, Hendrick J, Pijls M, Knape J. How frequently should basic cardiopulmonary resuscitation training be repeated to maintain adequate skills? *BMJ* 1983;306:1576-1577.
67. Jewkes F, Phillips B. Resuscitation training of paediatricians. *Arch Dis Child* 2003;88:118-121.
68. McKenna S, Glendon A. Occupational first aid training: decay in cardiopulmonary resuscitation (CPR) skills. *Journal of occupational psychology* 1985;58:109-117.
69. O'Hare B, Nakakeeto M, Southall D. A pilot study to determine if nurses trained in basic neonatal resuscitation would impact the outcome of neonates delivered in Kampala, Uganda. *J Trop Pediatr* 2006;52:376-379.
70. Wiswell T, Gannon C, Jacob J, Goldsmith L, Szyld E, Weiss K, et al. Delivery room management of the apparently vigorous meconium-stained neonate: results of the multicentre, international collaborative trial. *Pediatrics* 2000;105:1-7.
71. Carlo W, McClure E, Chomba E, Chakraborty H, Hartwell T, Harris H, et al. Newborn care training of midwives and neonatal and perinatal mortality rates in a developing country. *Pediatrics* 2010;126(5):e1064-71.
72. Rowe A, Rowe S, Holloway K, Ivanovska I, Muhe L, Lambrechts T. A systematic review of the effectiveness of shortening Integrated Management of Childhood Illness guidelines training: final report. World Health Organization. 2008.
73. Calderón C, Rotaecche R, Etxebarria A, Marzo M, Rico R, Barandiaran M. Gaining insight into the clinical practice guideline development processes: qualitative study in a workshop to implement the GRADE proposal in Spain. *BMC Health Serv Res* 2006(6):138.
74. Hirsh J, Guyatt G. Clinical experts or methodologists to write clinical guidelines? *Lancet* 2009;374:273-5.
75. Duke T, Campbell H, Ayieko P, Opiyo N, English M, Kelly J, et al. Accessing and understanding the evidence. *Bull World Health Organ* 2006;84(12):922.

76. Agency for Healthcare Research and Quality (AHRQ). <http://www.ahrq.gov/> (Accessed October 10, 2011).
77. Ayieko P, Ntoburi S, Wagai J, Opondo C, Opiyo N, Migiro S, et al. A multifaceted intervention to implement guidelines and improve admission paediatric care in Kenyan district hospitals: a cluster randomised trial. *PLoS Med* 2011;8(4):e1001018.
78. Gathara D, Opiyo N, Wagai J, Ntoburi S, Ayieko P, Opondo C, et al. Quality of hospital care for sick newborns and severely malnourished children in Kenya: a two-year descriptive study in 8 hospitals. *BMC Health Serv Res* 2011;11:307.
79. Grimshaw J, Eccles M, Tetroe J. Implementing clinical guidelines: current evidence and future implications. *J Contin Educ Health Prof* 2004;24(Suppl1:S31-7).
80. Graham W, Wagaarachchi P, Penney G, McCaw-Binns A, Antwi K, Hall M. Criteria for clinical audit of the quality of hospital-based obstetric care in developing countries. *Bull World Health Organ* 2000;78(5):614-20.
81. Wagaarachchi P, Graham W, Penney G, McCaw-Binns A, Yeboah Antwi K, Hall M. Holding up a mirror: changing obstetric practice through criterion-based clinical audit in developing countries. *Int J Gynaecol Obstet* 2001;74(2):119-30.
82. Kane M. The assessment of professional competence. *Eval Health Prof* 1992;15:163-182.
83. Gwer S, Newton C, Berkley J. Over-diagnosis and co-morbidity of severe malaria in African children: a guide for clinicians. *Am J Trop Med Hyg* 2007;77(6 Suppl):6-13.
84. Talbert A, Thuo N, Karisa J, Chesaro C, Ohuma E, Ignas J, et al. Diarrhoea complicating severe acute malnutrition in Kenyan children: a prospective descriptive study of risk factors and outcome. *PLoS One* 2012;7(6):e38321.
85. Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD. Audit and feedback: effects on professional practice and health care outcomes. *Cochrane Database of Syst Rev* 2006, Issue 2. Art. No.: CD000259. DOI: 10.1002/14651858.CD000259.pub2.
86. Grimshaw J, Thomas RE, MacLennan G, Fraser C, Ramsay CR, Vale L, et al. Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technol Assess* 2004;8(6)(iii-iv, 1-72).
87. O'Brien M, Rogers S, Jamtvedt G, Oxman AD, Odgaard-Jensen J, Kristoffersen DT, et al. Educational outreach visits: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2007, Issue 4. Art. No.: CD000409. DOI: 10.1002/14651858.CD000409.pub2.

Papers I-V

What clinical signs best identify severe illness in young infants aged 0–59 days in developing countries? A systematic review

Downloaded from adc.bmj.com on January 15, 2011 - Published by group.bmj.com
 Newton Opiyo,¹ Mike English^{1,2}

¹KEMRI/Wellcome Trust Research Programme, Nairobi, Kenya

²Department of Paediatrics, University of Oxford, Oxford, UK

Correspondence to
 Newton Opiyo, KEMRI/Wellcome Trust Research Programme, P.O. Box 43640, 00100 GPO, Nairobi, Kenya; nopiyo@nairobi.kemri-wellcome.org

Accepted 3 November 2010

ABSTRACT

Despite recent overall improvement in the survival of under-five children worldwide, mortality among young infants remains high, and accounts for an increasing proportion of child deaths in resource-poor settings. In such settings, clinical decisions for appropriate management of severely ill infants have to be made on the basis of presenting clinical signs, and with limited or no laboratory facilities. This review summarises the evidence from observational studies of clinical signs of severe illnesses in young infants aged 0–59 days, with a particular focus on defining a minimum set of best predictors of the need for hospital-level care. Available moderate to high quality evidence suggests that, among sick infants aged 0–59 days brought to a health facility, the following clinical signs—alone or in combination—are likely to be the most valuable in identifying infants at risk of severe illness warranting hospital-level care: *history of feeding difficulty, history of convulsions, temperature (axillary) $\geq 37.5^{\circ}\text{C}$ or $< 35.5^{\circ}\text{C}$, change in level of activity, fast breathing/respiratory rate ≥ 60 breaths per minute, severe chest indrawing, grunting and cyanosis.*

INTRODUCTION

Despite recent overall improvement in the under-five mortality worldwide, young infant mortality remains a serious problem, accounting for an increasing proportion of child deaths in resource-poor countries.^{1–2} Most young infant deaths continue to occur in homes with unwillingness, inability or delay in care seeking precluding appropriate referral of severely ill infants to adequately resourced health facilities.³ When healthcare is sought primary and even secondary health facilities (rural hospitals) in resource-poor countries often have no specialists (such as paediatricians) and limited or no laboratory diagnostic capability.⁴ In such settings, clinical decisions for appropriate management of severely ill infants have to be made on the basis of presenting clinical signs and symptoms alone. Typically health workers providing immediate care in these settings (even non-specialist physicians) have had as little as 2–3 weeks instruction in the care of the sick newborn in basic training courses lasting 2–5 years.

So which clinical symptoms and signs are the most useful in such settings for identifying serious illness in this vulnerable group of patients? The current Kenyan adaptation of the World Health Organisation (WHO) Integrated Management of Childhood Illness (IMCI) algorithm recommends a panel of 15 clinical signs and symptoms for the identification of possible severe disease in infants

aged 0–59 days (ie, young infants).⁵ The current panel of signs was based on the WHO multicentre study of clinical features and causes of serious bacterial infections in young infants.⁶ Training health workers to identify large numbers of signs and then using an algorithm based on all these signs in often busy clinics in resource-poor settings may threaten feasibility of implementation. We therefore sought to summarise the evidence available on clinical predictors of serious illnesses to help define a likely minimum set of signs that would be most useful in revised Kenyan national guidelines for the hospital care component of IMCI named Emergency Triage, Assessment, and Treatment plus Admission Care (ETAT+)^{7–8} and potentially to broader child survival programs such as the WHO's IMCI approach.

The clinical question addressed was: In sick young infants aged 0–59 days brought to a health-care worker, which clinical signs, alone or in combination, are most useful at indicating the presence of severe disease warranting referral-level care or hospitalisation for interventions that might include: parenteral antibiotics, parenteral fluids, assisted feeding, oxygen therapy, etc.? In particular, our interest was to identify a minimum set of clinical features that might best: (1) predict the need for treatment of potentially severe infection; (2) usefully limit the number and variety of clinical indicators health workers must be aware of that would comprise a basic, minimum standard for knowledge, clinical assessment and management; (3) help identify ill young infants for more specialist review if this is available.

Our interest was not therefore to identify all the clinical symptoms and signs that may be associated with serious illness in those aged 0–59 days. Rather the emphasis is on those signs and symptoms which most efficiently and effectively identify young infants at risk of severe disease after excluding those with prematurity, very low birth-weight or severe jaundice. We reasoned that such a minimum set of signs and symptoms should form the basis of practice, in managing possible neonatal sepsis in particular, for those with limited training or experience in young infant care if more specialist review is not available.

METHODS

Search strategy and selection criteria

Potential articles for inclusion were identified by direct searches of The Cochrane Library and MEDLINE (both from inception to November 2009). MEDLINE was searched via PubMed

clinical query filters. The searches were performed by combining MeSH (Medical Subject Headings) terms that are indicative of acute illnesses of interest (sepsis, bacteraemia), predictive of illness severity (signs, symptoms, clinical predictors, clinical markers) and indicative of target age group (neonates, infants and children). Further published and unpublished papers were sought by screening through bibliographies of identified articles and writing to authors of identified relevant papers. No language or time limits were applied in the search strategy. The complete search strategy is available from the authors.

Studies were included if they reported a set of clinical signs predictive of severe illnesses or mortality in young infants aged 0–59 days. Studies that included children aged 60 days and above were also considered if they also reported outcomes for children aged less than 60 days. Studies that examined a single disease state such as pneumonia or meningitis were not included in this review as arguably the clinical diagnosis of such specific diagnoses represents a more specialist task. Studies on bio-markers of severe illnesses (eg, C-reactive proteins) were excluded given the limited laboratory capability for their measurement in resource-poor settings. Studies conducted in high-income countries (as defined by the World Bank⁹) were excluded given the different spectrum and prevalence of severe illnesses. Only prospective studies with consecutive patient recruitments were considered. While a variety of definitions of ‘severe illness episodes’ have been suggested, in this paper severe illnesses were defined as cases warranting referral- or hospital-level care. Both community- and outpatient-based prospective observational studies were considered. Two reviewers independently screened through the titles and abstracts of identified articles, and applied the pre-defined selection criteria to assess their eligibility. Disagreements were resolved by discussion.

Assessment of quality of evidence

The strength of evidence—reflecting the appropriateness of the study design to answer the clinical question, the plausibility of prediction based on clinical signs, and the quality, quantity, and consistency of evidence—was independently assessed using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach.¹⁰ The approach classifies the quality of evidence (ie, ‘the extent to which one can be confident that an estimate of effect or association is correct’) into four categories: high, moderate, low, or very low (table 1). The unique features of GRADE include: (1) explicit, comprehensive criteria for downgrading and upgrading quality of evidence ratings; (2) explicit evaluation of the importance of outcomes and (3) clear separation of quality of evidence from the strength of recommendations. The GRADE evidence profiles were prepared by one reviewer (NO) and verified independently by a second reviewer (ME). Discrepancies in the quality ratings were resolved by discussion.

RESULTS

Study characteristics

Overall five^{4 6 11–13} prospective observational studies (n = 17 506 infants) out of 404 identified papers were included in this review (figure 1). All the included studies were conducted in resource-poor settings: three were based in outpatient clinics of first referral-level health facilities (basic or rural hospitals),^{6 12 13} one was community-based¹¹ while in another both outpatient and inpatient illness episodes were

considered.⁴ Three studies^{4 6 13} evaluated clinical predictors of severe illnesses while the remaining two^{11, 12} reported risk factors for death. The mean duration of recruitment was 12 months. The characteristics of the included studies are summarised in table 2.

In three studies,^{4 6 13} expert paediatrician opinion backed up with laboratory data (eg, blood or cerebrospinal fluid culture, chest radiography, pulse oximetry) was used as the diagnostic reference standard for severe illness classifications. There was however no ‘gold standard’ diagnostic reference in the remaining two studies, and the reported clinical signs were evaluated against a mortality outcome assessed by a neonatologist¹¹ or from review of primary healthcare workers¹² history taking and clinical examination. The quality of evidence for the suggested restricted set of best clinical predictors of severe illnesses was moderate to high (table 1).

Deriving the panel of best clinical predictors

A variety of statistical approaches were used in the individual reports reviewed to derive the best clinical predictors of severe illnesses and their combination (table 3). However, all used multivariable logistic regression models to adjust for known confounders (eg, place of study, age and weight) in attempts to improve the internal validity of the results. However, additional potential confounders and suppressors—such as differences in clinician practice, referral care patterns, prevalence of severe illnesses, HIV, patient case-mix, or temporal changes—were not adjusted for, and may further influence the performance of diagnostic algorithms in routine clinical settings. None of the derived sets of clinical signs reported in the studies identified has been the subject of further research to provide external validation or confirm effectiveness as has recently been recommended as appropriate for such an area of work.¹⁴

Comparability of individual study results is limited by the varied internal validation approaches used to develop the diagnostic algorithms: in three studies^{4 6 13} cross-validation was performed by re-calculating sensitivities and specificities (with 95% CIs) following omission of signs, one at a time, from an existing set of independent clinical predictors. In another study,¹¹ the performance of the derived set of ‘any two of seven signs’ (derivation set) was tested on a second ‘postintervention’ dataset (confirmatory set). The partial adjustment for potential confounders and the varied validation methods could increase the chances for observing heterogeneous sets of clinical signs. However, it should be noted that the two largest studies^{6 13} derived predictors from multi-country data and in one,¹³ country-specific results were also reported.

Clinical predictors of severe illnesses or death

Outlined below are results of the studies included that attempted to identify those signs that performed best, as a set, in terms of sensitivity and specificity (efficiency) for identifying severe neonatal and young infant illness.

In the largest ever study, a WHO multi-centre (Bangladesh, India, Pakistan, Bolivia, South Africa, Ghana) study¹³ on the clinical predictors of severe illnesses in children, 3177 neonates aged 0–6 days and 5712 infants aged 7–59 days brought with acute illnesses to health facilities were enrolled. Sepsis, pneumonia and meningitis were the most common diagnoses requiring hospital admission in both age groups, according to the gold-standard opinion, while those with severe jaundice were specifically excluded. A single algorithm (based on

Table 1 GRADE summary combining quality of evidence and summary of findings***Question:** What clinical signs best identify severe illness in young infants aged 0–59 days?**Settings:** Primary healthcare settings in resource-poor settings**Diagnostic criteria:** Clinical signs (clinical referral algorithms)

Quality assessment							Summary of findings		
No of studies	No of infants	Design	Limitations	Inconsistency	Indirectness	Imprecision	ORst (Range)	Quality (GRADE)	Importance
Cyanosis§ 3 ⁴ 6 ¹³	13 428	Observational studies	No serious limitations	No serious inconsistency	No serious indirectness	No serious imprecision	1.5–25.8	⊕⊕⊕⊕ HIGH	Critical
Change in level of activity‡ 3 ⁶ 11 ¹³	15 759	Observational studies	No serious limitations	No serious inconsistency	No serious indirectness	No serious imprecision	1.5–15.1	⊕⊕⊕⊕ HIGH	Critical
Fast breathing (respiratory rate ≥60 bpm) 3 ⁴ 6 ¹³	13 428	Observational studies	No serious limitations	No serious inconsistency	No serious indirectness	No serious imprecision	1.5–3.1	⊕⊕⊕⊖ MODERATE	Critical
Grunting 2 ⁶ 13	12 192	Observational studies	No serious limitations	No serious inconsistency	No serious indirectness	No serious imprecision	1.5–2.9	⊕⊕⊕⊖ MODERATE	Critical
History of convulsions 2 ⁶ 13	12 192	Observational studies	No serious limitations	No serious inconsistency	No serious indirectness	No serious imprecision	1.5–15.4	⊕⊕⊕⊕ HIGH	Critical
History of difficulty feeding 3 ⁴ 6 ¹³	13 428	Observational studies	No serious limitations	No serious inconsistency	No serious indirectness	No serious imprecision	1.5–10.0	⊕⊕⊕⊕ HIGH	Critical
Severe chest indrawing 4 ⁴ 6 ¹² 13	13 939	Observational studies	No serious limitations	No serious inconsistency	No serious indirectness	No serious imprecision	1.5–8.9	⊕⊕⊕⊖ MODERATE	Critical
Temperature (axillary) ≥37.5°C or <35.5°C 3 ⁴ 6 ¹³	13 428	Observational studies	No serious limitations	No serious inconsistency	No serious indirectness	No serious imprecision	1.5–9.2	⊕⊕⊕⊕ HIGH	Critical

*Quality of evidence—the extent to which we can be confident that an estimate of effect or association is correct. The judgements are based on the: study design (randomised vs observational studies); likelihood of bias; consistency of the results across the studies; precision (wide or narrow CIs) of overall estimates and; directness of the evidence with respect to the populations, interventions and settings where the proposed intervention may be used

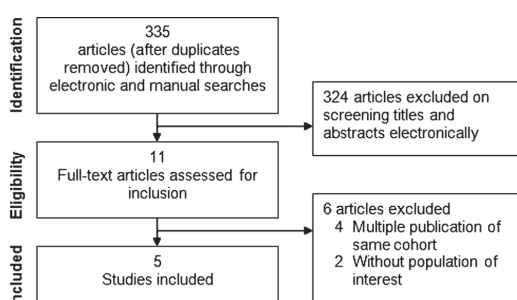
†ORs of signs or symptoms calculated by multivariable analyses

‡History of reduced activity, showing no spontaneous movement, stiff limbs, limbs becoming limp

§Bluish or greyish discoloration of the tongue

Quality of evidence is categorised as 'high', 'moderate', 'low' or 'very low'

- **HIGH:** Further research is very unlikely to change our confidence in the estimate of effect.
- **MODERATE:** Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
- **LOW:** Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
- **VERY LOW:** We are very uncertain about the estimate.

**Figure 1** Flow diagram of the study selection process

the prevalence of any one sign or symptom) of seven signs—*history of difficulty feeding, history of convulsions, movement only when stimulated, respiratory rate ≥60 breaths per minute (bpm), severe chest indrawing, temperature ≥37.5°C or <35.5°C*—had a sensitivity of 85% and a specificity of 75% in neonates aged 0–6 days. The 7 signs also did relatively well in infants aged 7–59 days (sensitivity 74%, specificity 75%). The authors suggested that this referral decision algorithm could be used

to predict the need for hospitalisation in all infants under 60 days of age who present to health facilities with acute illnesses.

In one Kenyan study⁴ of 1236 ill infants less than 60 days presenting to a rural district hospital, the presence of at least one of the following signs was 94% sensitive and 40% specific for severe disease (pneumonia, meningitis, prematurity, sepsis, acute respiratory infections, skin infections, purulent conjunctivitis) in infants aged 0–6 days: *a history of feeding difficulty, breathing difficulty, cough or abnormal behaviour, fever or indrawing*. In infants aged 7–59 days, the presence of at least one of the following signs was 97% sensitive and 56% specific for very severe disease: *a history of feeding difficulty, abnormal behaviour, breathing difficulty, fast breathing or indrawing, cyanosis and a bulging fontanelle*.

A re-analysis of an earlier WHO multicentre multi-country (Ethiopia, The Gambia, Papua New Guinea, The Philippines) study⁶ (n = 3303 infants) found the following clinical signs to be significantly associated (sensitivity 87%, specificity 54%) with severe disease in young infants less than 60 days old presenting with bacterial infections (pneumonia, hypoxaemia, bacteraemia, meningitis) at hospitals or outpatient clinics: *reduced feeding ability, no spontaneous movement, temperature >38.0°C, being drowsy or unconscious, a history of feeding problem*

Table 2 Characteristics of included studies

Study	Design	Setting; Country	Number of infants	Age range	Inclusion criteria	Exclusion criteria	Diagnostic reference standard	Common diagnoses	Mortality/1000
YICSSG ¹³	Prospective, consecutive	Outpatient Bangladesh, India, Pakistan, Bolivia, South Africa, Ghana	0–6 days: 3177 7–59 days: 5712	<60 days	Infants <60 days brought to the hospital or outpatient clinic with an acute illness	Well baby visits, non-resident in study area, previous enrolment in study, repeat episode of same illness Need for immediate cardiopulmonary resuscitation, hospitalisation in the previous 2 weeks (except for delivery), congenital malformation	Expert paediatrician backed up with laboratory data (eg, blood, CSF, chest radiography, pulse oximetry)	Sepsis, Pneumonia, Meningitis	0–6 days: Range: 0–120 7 to 60 days: Range: 0–70
English <i>et al</i> ⁴	Prospective, consecutive	Outpatient / Inpatient Kenya	1236	<60 days	Hospital-based birth cohort weighing ≥ 1.5 kg All infants aged <90 days admitted to hospital	Not reported	Admitting doctor backed up with simple diagnostic tests (full blood count, blood culture, lumbar puncture, chest x-ray)	Pneumonia, Meningitis, Prematurity, Sepsis, Acute respiratory infections, Skin infections, Purulent conjunctivitis	0–6 days: 320 7–60 days: 70
Weber <i>et al</i> ⁶	Prospective, consecutive	Hospitals or outpatient clinics Ethiopia, The Gambia, Papua New Guinea, The Philippines	3303	<60 days	Infants <91 days with possible acute infections	Infants with congenital heart disease and hypoxemia	Expert paediatrician backed up with laboratory data (eg, blood culture, chest radiograph, lumbar puncture, pulse oximetry)	Pneumonia, Hypoxaemia, Bacteremia, Meningitis	0–59 days: 59
Dukeet <i>et al</i> ¹²	Prospective, consecutive	Outpatient clinic Papua New Guinea	511	<60 days	All sick young infants presenting to rural health centres	Not reported	Mortality outcome following a review of primary healthcare workers' history taking and clinical examination	Neonatal sepsis, Pneumonia, Malaria	0–59 days: 59.8
Banget <i>et al</i> ¹¹	Prospective, consecutive	Community India	3567	<28 days	All neonates born in study villages	Not reported	Mortality outcome judged by a neonatologist as due to sepsis	'Sepsis' (defined as sepsis, meningitis, pneumonia)	0–28 days: 51.2

CSF, cerebrospinal fluid, YICSSG, Young Infants Clinical Signs Study Group.

or change in activity, agitation, lower chest wall indrawing, respiratory rate >60 bpm, grunting, cyanosis, convulsions, bulging fontanelle and slow digital capillary refill.

In one multi-site (single country) study¹² of 511 infants less than 60 days of age presenting to rural health centres in Papua New Guinea, the following signs were associated with an increased risk of death: *inability to feed, fast respiratory rate (fast breathing), apnea, cyanosis, 'too small', 'skin-cold' and severe abdominal distension*. The most common diagnoses included neonatal sepsis, pneumonia and malaria. The authors concluded that the above signs could be used as triggers for emergency care, longer observation or urgent referral.

Finally, in one field study¹¹ of 3567 neonates aged less than 28 days in India, simultaneous presence of any two of the following seven clinical signs predicted death from sepsis with a 100% sensitivity and 92% specificity: *reduced or stopped sucking, weak or no cry, limbs becoming limp, vomiting or abdominal distension, baby cold to touch, severe chest indrawing and umbilical infection*. The authors concluded that these criteria can be used by health workers to select sick neonates for treatment or referral.

Taken together, and based on the overlap of study results and the consistency of performance of clinical symptoms and

signs in identifying severe illness, moderate to high quality evidence (table 1 and 4) suggest that the following eight clinical signs—based on their strengths of associations (ORs) with severe illnesses, prevalence in the enrolled infants in the primary studies, and ease of clinical recognition—are likely to be the most valuable in predicting severe illnesses in young infants presenting at primary healthcare facilities: *history of feeding difficulty, history of convulsions, temperature (axillary) $\geq 37.5^{\circ}\text{C}$ or $<35.5^{\circ}\text{C}$, change in level of activity, fast breathing/respiratory rate ≥ 60 bpm, severe chest indrawing, grunting and cyanosis*. A suggested more sensitive alternative to cyanosis, and which has been shown to be strongly associated with mortality would be hypoxaemia—economically and reliably diagnosed using pulse oximetry.^{15 16}

DISCUSSION

Interpretation of findings

This review set out to define a set of simple best clinical predictors of severe illnesses in infants aged 0–59 days. The limited set of clinical signs for which extensive evidence supporting their value exists were reported to have high sensitivity (indicating that they were less likely to miss severe illness

Table 3 Deriving best clinical predictors of severe illnesses

Study	Analytical strategy
Bang <i>et al</i> ¹¹	The sensitivity and specificity of 16 signs significantly associated with sepsis death (lower 95% CI of ORs>1 were calculated to identify a prediction rule of any two of a set of six signs. The best set of 'any two of seven criteria' (100% sensitivity, 92% specificity) was subsequently selected by sequentially adding signs of the respiratory system*, one at a time, to previously selected six signs.
Duke <i>et al</i> ¹²	Clinical signs associated with death were identified using univariate logistic regression ORs (<0.05). A panel of four independent clinical predictors of death (ORs 3.6 to 6.2) was identified by multivariate logistic regression analyses using independent predictors that were present in one third or more of the deaths.
English <i>et al</i> ⁴	Signs significantly associated with very severe illness were identified by univariate ORs—calculated using logistic regression that took account of likely collinearity† between signs. Panel of best clinical predictors (0–6 days: 94% sensitivity, 40% specificity; 7–59 days: 97% sensitivity, 56% specificity) was subsequently derived by subjecting the identified independent predictors to multiple multivariate logistic regressions.
Weber <i>et al</i> ⁶	A set of independent predictors (ORs>2.5) identified (by univariate logistic regression analyses) from an expert selected panel of candidate signs. The final set of best clinical predictors (87% sensitivity, 54% specificity) was subsequently derived from multivariate analyses of several combinations of independent predictors of severe illnesses.
YICSSG ¹³	A panel of 12 independent clinical predictors of the need for urgent hospital care was identified by univariate logistic regressions. A further reduction of the list to seven signs (ORs 2.7–15.4, p<0.05) was made on the basis of low prevalence of some signs and negligible change in sensitivity (calculated by random-effects meta-analysis‡) if they were omitted.

*Respiratory rate ≥60 breaths per minute, chest indrawing and grunting

†Signs with similar ORs

‡Weights studies more equally/yields more conservative estimates

YICSSG, Young Infants Clinical Signs Study Group.

episodes) and reasonable specificity (indicating that they were likely to reduce unnecessary hospitalisation or referral). These symptoms and signs were supported by data from large prospective observational studies conducted in resource-limited healthcare settings. Such settings are likely to reflect typical busy clinical situations in many resource-poor country health facilities with limited laboratory diagnostic facilities and high health worker workloads.

The eight identified symptoms and signs—history of feeding difficulty, history of convulsions, temperature (axillary) ≥37.5°C or <35.5°C, change in level of activity, fast breathing/respiratory rate ≥60 bpm, severe chest indrawing, grunting and cyanosis—are therefore probably the most appropriate to employ as a basic, minimum standard for knowledge, clinical assessment and management for health workers with limited training or experience in the care of sick newborns or young infants working in rural primary healthcare settings or emergency outpatient clinics of district hospitals in resource-poor countries. Presence of any one of these danger signs should prompt health workers with only basic training to initiate treatment for serious illness until an early opinion or review is available from a health worker with a higher level of training or experience.

Such an approach prioritises sensitivity (not missing a true serious illness) at the expense of specificity (restricting treatment of those without serious illness) in a population of vulnerable patients. Thus, it should be remembered that the presence of any one of the clinical danger signs does not provide a reliable clinical diagnosis but rather a reasonable basis for initiating empiric treatment. Such an approach is justified given the high mortality in the neonatal period (which has been documented to be 40% of all under-five child deaths globally, with 99% in resource-poor countries¹) and the limited training and skills of qualified health workers with only basic training who comprise the majority of those caring for patients even at hospital levels. It should however be remembered that these referral/empiric treatment criteria are not necessarily applicable to illness episodes for presentations with the primary problems of jaundice (since severity would depend on the level of hyperbilirubinemia), birth asphyxia or prematurity.

The focus of this review was specifically on identifying a minimum set of signs and symptoms that health workers should be able to identify with the goal of efficiently initiating empiric treatment or specialist referral. Clinical features that do not necessarily improve the sensitivity and specificity of this set may nonetheless be strongly associated with the outcome of serious illness (table 5). Although studies varied in the range of signs and symptoms examined, in univariate analyses, at least two studies indicated strong associations between serious illness and signs that are perhaps worthy of further evaluation. Pallor, slow capillary refilling and a bulging fontanelle may be indicative of either specific but uncommon pathologies and/or the need for specific interventions that go beyond standard empiric antibiotics, provision of oxygen and feeding support.

In this review, the use of the GRADE approach added scientific rigor to the process of compiling and rating the quality of evidence. Our experience suggests that it is feasible to use GRADE even for evaluations of diagnostic/screening interventions. However, a number of challenges remain, particularly regarding assessment of the: (1) range of baseline (control) risks (a useful measure of the typical burden of outcomes)—as these remain largely under-reported in diagnostic observational studies and; (2) likelihood of publication (reporting) bias given the heterogeneous reporting of diagnostic outcome data.

Limitations of summarised evidence

First, a limited number of well-conducted studies (N=5) were available for inclusion in this review and the heterogeneous nature of available outcome data made it impossible to statistically assess the influence of publication bias on the results. However, the five studies enrolled (consecutively) a large number of infants (N=17 506), and the current results would therefore be expected to be robust to inclusions of any un-retrieved eligible published or unpublished studies. Second, the main aim of clinical algorithms is to identify severe illness so that appropriate treatment is initiated promptly; clinical predictors of death (reported in two studies) may therefore be of limited value—as they indicate advanced stages of disease during which treatment may be less likely to work. Finally, the lack of prospective studies

Table 4 Independent clinical predictors of severe illness in young infants*

	Bang <i>et al</i> ^{†1}	Duke <i>et al</i> ^{†12}	English <i>et al</i> ⁴	English <i>et al</i> ⁴	Weber <i>et al</i> ⁶	YICSSG ^{‡13}	ORs (range)
	0–28 days	0–59 days	0–6 days	7–59 days	0–60 days	0–6 days	
Feeding							
History of difficult feeding	–	–	7.3 (3.1–16.8)	2.8 (2.6–5.0)	>1.5	10.0 (6.9–14.5)	1.5–10.0
Reduced feeding ability	–	SP	–	–	>1.5	–	1.5–7.4
Sucking weak, reduced or stopped	7.9 (1.8–34.2)	–	–	–	–	–	7.9
Activity							
History of change in level of activity	–	–	–	–	>1.5	–	1.5
Lethargy	–	–	–	–	–	3.5 (1.7–7.1)	3.5
Limbs becoming limp	3.3 (0.9–12.0)	–	–	–	–	–	3.3
Movement only when stimulated	–	–	–	–	–	6.9 (3.0–15.5)	6.9
No spontaneous movement	–	–	–	–	>1.5	–	1.5
Stiff limbs	–	–	–	–	–	15.1 (2.2–105.9)	15.1
Respiratory							
Apnea	–	4.2 (1.1–15.4)	–	–	–	–	4.2
Cough	–	–	0.1 (0.02–0.5)	–	–	–	0.1
Difficulty breathing	–	–	2.1 (1.0–2.6)	1.8 (0.9–3.5)	–	–	1.8–2.1
Fast breathing	–	–	–	3.1 (1.8–5.3)	–	–	3.1
Grunting	–	–	–	–	>1.5	2.9 (1.1–7.5)	1.5–2.9
Severe (deep) lower chest indrawing	–	3.6 (0.94–13.9)	3.0 (1.1–8.2)	2.4 (1.3–4.7)	>1.5	8.9 (4.0–20.1)	1.5–8.9
Respiratory rate ≥60 breaths per minute	–	–	–	–	>1.5	2.7 (1.9–3.8)	1.5–2.7
Skin							
Cyanosis	–	–	–	25.8 (1.9–354)	>1.5	13.7 (1.6–116.5)	1.5–13.7
Prolonged capillary refill	–	–	–	–	>1.5	10.5 (5.1–21.7)	1.5–10.5
'Skin cold' ('baby cold to touch')	3.5 (1.0–12.4)	6.2 (1.5–26.6)	–	–	–	–	3.5–6.2
Temperature <35.5°C	–	–	–	–	–	9.2 (4.6–18.6)	9.2
Temperature (axillary) ≥37.5°C	–	–	3.2 (1.7–6.3)	–	>1.5	3.4 (2.4–4.9)	1.5–3.4
Conscious state							
Conscious state agitated	–	–	–	–	>1.5	–	1.5
Unconscious or drowsy	–	–	–	–	>1.5	–	1.5
Others							
Abnormal behaviour	–	–	2.4 (1.2–4.6)	3.1 (1.7–5.6)	–	–	2.4–3.1
Bulging fontanelle	–	–	–	1.9 (3.0–39.9)	>1.5	–	1.5–1.9
Cry abnormal, weak, or stopped	14.3 (3.9–52.1)	–	–	–	–	–	14.3
History of convulsions	–	–	–	–	>1.5	15.4 (6.4–37.2)	15.4
Severe abdominal distension / vomiting	6.8 (1.7–27.2)	–	–	–	–	–	6.8

OR not reported; (p<0.001)

*Values are multivariate ORs with 95% CIs

†Studies reporting predictors (risk factors) for death

‡Panel of best clinical predictors had comparable sensitivities and specificities in 0–6 days and 7–59 days age groups

SP, significant predictor; YICSSG, Young Infants Clinical Signs Study Group.

confirming clinical effectiveness of the referral algorithms after implementation has also previously been noted.

In our narrative summary of study findings variation in the strengths of association (ORs) of specific signs and symptoms and serious illness was apparent. A possible explanation for the differences in the magnitude of ORs could be differences in the spectrum (case-mix or co-morbidities) and prevalence of illness episodes—for example, studies may vary in the proportion of cases that are meningitis or include a larger subset of severely ill infants (manifest as a higher death rate). Where populations studied vary, likelihood ratios¹⁷—which are more robust to changes in disease prevalence compared to sensitivity and specificity, and that were computed in only 1 study⁶—might have yielded more rigorous estimates of the association of clinical signs with severe illnesses. Another possible explanation for the differences in the ORs could be differences in the 'gold standard' criteria for verification of severe illness (ie, likelihood of 'reference standard misclassification')—for example, clinical signs such as cyanosis may be detected and interpreted more accurately in studies where the reference

standard was an experienced neonatologist assisted by pulse oximetry. These factors may explain the relatively higher ORs reported in the largest multi-centre study.¹³

None of the included studies considered the cost-benefit implications of reported clinical referral algorithms compared to alternative diagnostic strategies—such as illness severity scoring algorithms (eg, the Baby Check¹⁸) or rapid point-of-care laboratory bio-markers of illness severity (eg, C-reactive proteins¹⁹). However, use of the minimum subset of signs and symptoms represents a refinement of the currently larger set of fifteen signs recommended in Kenyan IMCI guidelines (table 6)⁵ and might be expected to be implemented more easily and efficiently. The suggested panel of eight signs of severe illness includes all the six signs recommended in the revised WHO IMCI guidelines (table 6).²⁰

Implications for practice and policy

The set of diagnostic features proposed as a basic algorithm for initiating referral/empiric treatment should be feasible to implement as part of revised IMCI strategy including those

aimed at first referral level facilities staffed by health workers with only basic training. To promote adherence and improve the diagnostic value of the algorithm, we suggest: (1) concurrent implementation and scaling up of community-based intervention strategies aimed at improving early healthcare seeking behaviour for *any* of the suggested best predictors of severe illnesses, for example, through danger-signs sensitisation-

health education for families¹⁹ and; (2) preservice and inservice (refresher) training for health workers on recognition and interpretation of the suggested danger signs (given their subtle nature of presentation and the low prevalence of severe young infant illnesses). Such measures should ideally be linked to efforts to improve empiric treatment, supportive care and access to healthcare providers with higher levels of training.

Table 5 Significant predictors* of severe illnesses in univariate analyses

	Bang <i>et al</i> ¹¹	Duke <i>et al</i> ¹²	English <i>et al</i> ⁴	English <i>et al</i> ⁴	Weber <i>et al</i> ⁶	YICSSG 2008† ¹³	YICSSG 2008† ¹³
	0–28 days	0–59 days	0–6 days	7–59 days	0–60 days	0–6 days	
Abdominal distension	–	–	–	–	–	4.6 (2.2–9.7)	–
Abnormal movements	–	–	4.5 (2.6–7.6)	3.3 (2.0–5.6)	–	–	–
Bulging fontanelle	–	18.7 (2.5–141.8)	2.8 (0.3–32.0)	–	–	5.6 (1.8–18.0)	9.6 (3.1–29.9)
Chest indrawing	–	3.7 (1.4–9.9)	–	–	–	–	–
Consolability: continues to cry/fuss	–	–	–	–	2.9 to 4.0‡	–	–
Cyanosis	–	–	2.4 (1.0–5.6)	–	–	35.0 (10.0–122.7)	7.1 (2.5–20.3)
Drowsy / unconscious	40.2 (14.0–116.6)	–	–	–	–	–	–
Grunting	7.0 (2.6–18.7)	–	–	2.4 (1.3–4.8)	–	12.6 (4.1–38.7)	9.7 (5.4–17.4)
History of blood in stool	–	–	–	–	–	–	6.1 (2.2–16.9)
History of change in crying	–	–	–	–	1.9 (1.4–2.7)	–	–
History of cough	–	–	–	–	1.5 (1.1–2.0)	–	–
History of diarrhoea	–	–	–	–	–	–	1.5 (1.1–2.2)
History of fever	–	–	–	–	–	2.4 (1.8–3.2)	2.9 (2.3–3.7)
History of no cry at birth	–	–	–	–	–	2.6 (1.3–5.4)	–
Hypothermia†	–	5.3 (1.5–18.8)	–	–	–	–	–
Lethargic	–	–	3.1 (1.8–5.3)	2.6 (1.5–4.3)	–	20.5 (13.8–30.5)	24.0 (15.6–36.9)
Nasal flaring	–	–	2.2 (1.0–4.9)	3.0 (1.9–4.8)	–	15.7 (5.7–43.1)	14.8 (7.8–28.2)
Pallor	–	37.5 (3.2–436.8)	2.8 (0.3–31.8)	16.4 (3.4–78.0)	–	–	–
Prolonged capillary refill	–	–	–	–	–	12.1 (5.2–28.3)	31.6 (11.8–84.3)
Reduced skin turgor	–	–	–	–	–	3.7 (2.2–6.2)	15.7 (6.6–37.4)
Restless and irritable	–	–	–	–	–	7.2 (2.4–21.3)	13.9 (6.8–28.3)
Stiff limbs	–	–	–	–	–	44.9 (10.7–188.2)	7.8 (2.1–29.2)
Sunken eyes	–	–	–	–	–	–	11.5 (3.7–35.6)
Temperature (axillary) 35.0°C	11.5 (4.5–30.0)	–	–	–	–	–	–
Unconscious	–	–	5.0 (1.7–14.3)	3.9 (1.6–9.6)	–	–	–

*Signs significantly associated with severe illnesses (p values <0.05) not included in final multivariable models; Numbers are univariate ORs with 95% CIs

†Axillary temperature <36.0°C

‡Range of ORs for association with severe disease (sepsis, meningitis or hypoxemia)

YICSSG - Young Infants Clinical Signs Study Group.

Table 6 Comparison of current Kenyan IMCI referral criteria, revised WHO criteria and proposed criteria based on studies included in this review

Current Kenyan IMCI referral criteria ⁵	Revised WHO IMCI referral criteria ²⁰	Proposed referral criteria based on studies included in this review ^{4,6,11–13}
Not able to feed or breastfeed	Not feeding well	History of feeding difficulty
Convulsions or convulsing now	Convulsions	History of convulsions
Fast breathing (≥60 bpm or more)	Fast breathing (≥60 bpm or more)	Fast breathing (respiratory rate ≥60 bpm)
Severe chest indrawing	Severe chest indrawing	Severe chest indrawing
Fever (≥37.5°C* or feels hot) or low body temperature (<35.5°C* or feels cold)	Fever (≥37.5°C*) or low body temperature (<35.5°C*)	Temperature (axillary) ≥37.5°C or <35.5°C
No movements even when stimulated	Movement only when stimulated or no movement at all	Change in level of activity
Grunting or wheezing	–	Grunting
Central cyanosis	–	Cyanosis
Gasping	–	–
Not breathing at all even when stimulated	–	–
Respiratory rate less than 20 bpm	–	–
Nasal flaring	–	–
Bulging fontanelle	–	–
Pus draining from the ear	–	–
Drowsy (lethargic) or unconscious	–	–

*Axillary temperature
bpm, breaths per minute.

Implications for future research

The findings of this review have a number of important implications for future research. First, we suggest further large observational validation studies to confirm the effectiveness of the proposed minimum set of eight clinical signs and symptoms in routine practice. Second, we suggest research should examine the approach's diagnostic performance among HIV-infected infants.²¹ Finally, improvements to this clinical approach, possibly by combining it with bio-markers of severe illness, should be examined.

CONCLUSION

The findings of this review suggest that, among sick infants aged 0–59 days, brought to a healthcare worker with only basic training, the following clinical signs—alone or in combination—indicate severe illness warranting referral or hospitalisation and empiric treatment in the absence of a senior opinion: *history of feeding difficulty, history of convulsions, temperature (axillary) $\geq 37.5^{\circ}\text{C}$ or $< 35.5^{\circ}\text{C}$, change in level of activity, fast breathing/respiratory rate ≥ 60 bpm, severe chest indrawing, grunting, and cyanosis*. Focusing only on health worker triaging skills without addressing the barriers to healthcare seeking may limit the impact on mortality of the referral algorithm—hence the need for concurrent implementation of interventions to improve care-seeking.

Acknowledgements This manuscript is published with the permission of the Director of KEMRI.

Funding Mike English is funded by a Wellcome Trust Senior Fellowship (#076827) and Newton Opiyo is supported by funding from a Wellcome Trust Strategic Award (#084538).

Competing interests None.

Contributors ME conceived the idea for the review. NO and ME screened records for eligibility, assessed quality of included studies and interpreted findings. NO prepared the first draft of the review.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

- Central Bureau of Statistics and Ministry of Planning and National Development. Kenya Demographic and Health Survey - Preliminary Report. Nairobi: Government of Kenya, 2008.
- Joy L, Kate K. *Opportunities for Africa's Newborns. Practical Data, Policy and Programmatic Support for Newborn Care in Africa*. Geneva, Switzerland: The Partnership for Maternal, Newborn and Child Health 2006.
- World Health Organisation. The World Health Report 2005: Make Every Mother and Child Count. Geneva, Switzerland: WHO, 2005.
- English M, Ngama M, Mwalekwa L, et al. Signs of illness in Kenyan infants aged less than 60 days. *Bull World Health Organ* 2004;**82**:323–9.
- Ministry of Health. Integrated Management of Childhood Illness. Kenya: Ministry of Health, 2008.
- Weber M, Carlin J, Gatchalian S, et al. Predictors of neonatal sepsis in developing countries. *Pediatr Infect Dis J* 2003;**22**:711–16.
- Irimu G, Wamae A, Wasunna A, et al. Developing and introducing evidence based clinical practice guidelines for serious illness in Kenya. *Arch Dis Child* 2008;**93**:799–804.
- Ministry of Medical Services. Basic Paediatric Protocols - Revised. Kenya: Government of Kenya, 2010.
- World Bank 2006. Data and Statistics, World Bank. <http://web.worldbank.org/> (accessed 4 December 2009).
- Schünemann H, Oxman AD, Brozek J, et al. GRADE Working Group. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* 2008; **336**:1106–10.
- Bang AT, Bang RA, Reddy MH, et al. Simple clinical criteria to identify sepsis or pneumonia in neonates in the community needing treatment or referral. *Pediatr Infect Dis J* 2005;**24**:335–41.
- Duke T, Oa O, Mokela D, et al. The management of sick young infants at primary health centres in a rural developing country. *Arch Dis Child* 2005;**90**:200–5.
- Young Infants Clinical Signs Study Group. Clinical signs that predict severe illness in children under age 2 months: a multicentre study. *Lancet* 2008;**371**:97–8.
- Altman DG, Vergouwe Y, Royston P, et al. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009;**338**:b605.
- Mwaniki MK, Nokes DJ, Ignas J, et al. Emergency triage assessment for hypoxaemia in neonates and young children in a Kenyan hospital: an observational study. *Bull World Health Organ* 2009;**87**:263–70.
- Subhi R, Adamson M, Campbell H, et al. The prevalence of hypoxaemia among ill children in developing countries: a systematic review. *Lancet Infect Dis* 2009;**9**:219–27.
- Centre for Evidence-Based Medicine. Likelihood ratios. <http://www.cebm.net/index.aspx?o=1043> (accessed 18 January 2010).
- Morley CJ, Rashid H, Thomas IR, et al. Use of Baby Check to assess the severity of illness in babies attending a clinic in The Gambia. *J Trop Pediatr* 1994;**40**:144–8.
- Bazzano AN, Kirkwood BR, Tawiah-Agyemang C, et al. Beyond symptom recognition: care-seeking for ill newborns in rural Ghana. *Trop Med Int Health* 2008; **13**:123–8.
- WHO IMCI chart booklet. http://www.who.int/child_adolescent_health/documents/IMCI_chartbooklet/en/ (accessed 12 October 2010).
- Carroll E, Mankambo LA, Jeffers G, et al. The diagnostic and prognostic accuracy of five markers of serious bacterial infection in Malawian children with signs of severe illness. *Arch Dis Child* 2009;**94**:660.

Effect of Newborn Resuscitation Training on Health Worker Practices in Pumwani Hospital, Kenya

Newton Opiyo^{1*}, Fred Were², Fridah Govedi³, Greg Fegan^{1,4}, Aggrey Wasunna², Mike English^{1,5}

1 Kenya Medical Research Institute (KEMRI)/Wellcome Trust Research Programme, Nairobi, Kenya, **2** Department of Paediatrics and Child Health, College of Health Sciences, University of Nairobi, Nairobi, Kenya, **3** Pumwani Maternity Hospital, Nairobi, Kenya, **4** Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom, **5** Department of Paediatrics, University of Oxford, Oxford, United Kingdom

Abstract

Background: Birth asphyxia kills 0.7 to 1.6 million newborns a year globally with 99% of deaths in developing countries. Effective newborn resuscitation could reduce this burden of disease but the training of health-care providers in low income settings is often outdated. Our aim was to determine if a simple one day newborn resuscitation training (NRT) alters health worker resuscitation practices in a public hospital setting in Kenya.

Methods/Principal Findings: We conducted a randomised, controlled trial with health workers receiving early training with NRT (n = 28) or late training (the control group, n = 55). The training was adapted locally from the approach of the UK Resuscitation Council. The primary outcome was the proportion of appropriate initial resuscitation steps with the frequency of inappropriate practices as a secondary outcome. Data were collected on 97 and 115 resuscitation episodes over 7 weeks after early training in the intervention and control groups respectively. Trained providers demonstrated a higher proportion of adequate initial resuscitation steps compared to the control group (trained 66% vs control 27%; risk ratio 2.45, [95% CI 1.75–3.42], $p < 0.001$, adjusted for clustering). In addition, there was a statistically significant reduction in the frequency of inappropriate and potentially harmful practices per resuscitation in the trained group (trained 0.53 vs control 0.92; mean difference 0.40, [95% CI 0.13–0.66], $p = 0.004$).

Conclusions/Significance: Implementation of a simple, one day newborn resuscitation training can be followed immediately by significant improvement in health workers' practices. However, evidence of the effects on long term performance or clinical outcomes can only be established by larger cluster randomised trials.

Trial Registration: Controlled-Trials.com ISRCTN92218092

Citation: Opiyo N, Were F, Govedi F, Fegan G, Wasunna A, et al (2008) Effect of Newborn Resuscitation Training on Health Worker Practices in Pumwani Hospital, Kenya. PLoS ONE 3(2): e1599. doi:10.1371/journal.pone.0001599

Editor: Jose M. Belizan, Institute of Clinical Effectiveness and Health Policy, Argentina

Received: October 1, 2007; **Accepted:** December 20, 2007; **Published:** February 13, 2008

Copyright: © 2008 Opiyo et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was supported in part by a grant from the Laerdal Foundation of Acute Medicine, Norway and in part by a Wellcome Trust Senior Research Fellowship awarded to Mike English (#076827). The Laerdal foundation had no role in the design or conduct of the study, made no contribution to the data analysis nor report writing and their permission to publish the results was not sought.

Competing Interests: The authors have declared that no competing interests exist.

*E-mail: nopiyo@nairobi.kemri-wellcome.org

Introduction

Birth asphyxia is estimated to cause 0.7 to 1.6 million deaths a year globally with 99% of these deaths occurring in developing countries [1]. Effective resuscitation could prevent some of these deaths as well as improve the outcomes of surviving asphyxiated babies [1]. However, provision of appropriate newborn resuscitation care is dependent on the presence of an adequately skilled health worker in the home or the facility. To date little attention has been paid to furnishing health workers with these skills and we have little idea what works. We do however know that inappropriate, ineffective or dangerous forms of practice are widespread [2,3,4].

In higher income settings Newborn Life Support (NLS) training courses have proliferated. Although these can be expensive little is known about the effect they actually have on health worker behaviour in practice settings [5]. Where studies on the effect of life support training for any age group have been done they focus mostly on knowledge and skill retention observed in simulated practice following course participation. Few studies have examined

outcomes considered more useful such as morbidity, mortality or work-place provider practices [5]. Furthermore, the few studies on provider behaviour were all methodologically weak and therefore very little confidence could be attached to their results [5]. The aim of this study was therefore to determine if a simple, one day newborn resuscitation training alters health worker resuscitation practices in a busy public hospital in a low-income setting.

Methods

The protocol for this trial and supporting CONSORT checklist are available as supporting information; see Checklist S1 and Protocol S1.

Participants and Randomisation Procedure

The study was conducted in Pumwani Maternity Hospital in Nairobi, Kenya. This is the main maternity facility for Nairobi and provides delivery care to 17,000 women each year. The hospital has approximately 90 nurse/midwives (60 assigned to the labour

ward and 30 to the theatre) primarily responsible for delivery care and newborn resuscitation with 14 on duty at any one time (8 labour ward, 6 theatre). A 150 bed newborn nursery, supervised by two paediatricians, provides care for all infants requiring medical attention after delivery. The labour ward has 8 cubicles where deliveries are conducted with resuscitations being performed on one resuscitaire, located no more than 10 metres from the furthest room. The theatre has 2 operating rooms each with a resuscitaire.

Our intention was to test resuscitation training on practices by randomly assigning labour ward and theatre staff to either early or late training, considering the health worker as a unit of clustering. Potential participants, the 90 nurse / midwifery staff, were therefore initially listed by place of work. Eligibility criteria for initial randomisation were: personal work plans for the 3 months post-training that neither included leave of >2 weeks duration, nor rotation to another work station; routine responsibility for newborn resuscitation; provision of informed consent. We aimed to ensure an equal proportion of staff (35%) from labour ward and theatre were included in the early training as this could accommodate at most 32 participants. Those not included in the early training were trained after the initial 3 months observation period.

Intervention

The intervention was purposely designed by the investigators together with representatives of the Kenya Resuscitation Council under the umbrella of the Kenya Paediatric Association. The form of training drew heavily on the one day UK Resuscitation Council training [6] in form but was significantly adapted to the Kenyan setting where resources are limited. The one day course teaches an A (Airway), B (Breathing) and C (Circulation) approach to resuscitation laying down a clear step by step strategy for the first minutes of resuscitation at birth. It comprises focused lectures aimed at understanding the modern approach to resuscitation and practical scenario sessions using infant manikins to develop skills in airway opening, use of a bag-valve-mask device and chest (cardiac) compressions. Candidates were provided with a simple instruction manual two weeks before the training for self-learning. At the end of the day trainees were assessed using a multiple-choice examination and a formal test scenario evaluating actual practical skills and their integration into a clinical context. Course instructors had completed a Kenya Resuscitation Council Advanced Life Support Generic Instructor Course (GIC) co-supervised by an experienced team from the UK resuscitation council.

Outcome Measures

The primary outcome for the study was the proportion of resuscitation episodes in which appropriate initial resuscitation steps were practiced as recommended in the NLS training. The primary outcome was further classified into two levels: perfect (where the health worker entirely followed the training guideline) and adequate resuscitation with minor, clinically insignificant deviations from the training guideline (see Appendix S1). The primary steps in recommended resuscitation include only the practices of: suction, restricted only to babies born through meconium yet to take a breath, drying (stimulating), airway examination (A) and positioning and assessment of breathing (B). These practices should occur within the first sixty seconds of any resuscitation making rapid assessment of correct practice possible for an observer. After this actions should depend on whether breathing and subsequently an adequate heart rate are detected, information not necessarily available to an observer. We therefore concentrated on the very early steps as our primary outcome because they should be universal, are readily observable and are

objective. In addition, if any problem is identified and the health worker calls for help then for ethical reasons the observers were instructed to provide whatever help they could, under instruction of the primary provider, only recording the step by step actions / instructions of the health worker as soon as possible thereafter. Secondary outcomes were: the frequency of inappropriate and/or potentially harmful practices which might confer a direct risk to the baby or an indirect risk through the delayed initiation of appropriate interventions (see Appendix S2); an overall score awarded to each resuscitation episode after independent review of the documented process by two NLS instructors blinded to the identity or training status of the health worker.

To capture data, trained observers worked a shift pattern to ensure at least one was present in the hospital continuously (spanning all 24 hours) until the estimated number of observations required by our sample size calculations were achieved. When two observers were available (approximately 30% of shifts) one remained on labour ward and one in theatre. When one observer was present they were assigned to either labour ward or theatre by one of the investigators (NO) who was aware of the training allocation to ensure that an adequate number of observations could be collected from each trained health worker. Resuscitation observers were nursing students who had been specially trained as a group over 3 days to make structured observations on newborn resuscitation using role play and scenarios and a standardised checklist. They were not trained in newborn life support. The observers were blind to the training status of the health workers and were instructed not to try to ascertain health workers' training status after discussing with them the possible biases this might introduce and their role in producing a valid research result.

The practice observation check list was based on the resuscitation steps included in the training. Data on events preceding the resuscitation episode, the health workers record of the baby's APGAR score, the availability of equipment and the outcome of the resuscitation were also recorded. All health workers were assigned a unique study code that was the only identifier used on all observation forms.

Routine data on delivery outcomes, admissions to nursery and their causes and outcomes were collected retrospectively for the 6 months prior to the first training (June 2006), for a period of 3 months between the first training and training of the remaining staff (September 2006) and for 3 months after this. We refer to the period between early and late training allowing comparison of practices in trained and untrained providers as phase 1 of the study. In addition, we aimed to observe 50 consecutive resuscitation episodes after the late training to describe practices after 'saturation training', this period is referred to as phase 2 of the study.

Sample-Size Calculations

Our sample size calculation took into account the clustered nature of our data, i.e. resuscitations by the same health worker. Based on routine hospital practice we estimated at best that 3 to 5 observations could be made per health worker over a 6 to 7 week period, a period we reasoned was short enough to reduce the possible effects of cross-group contamination. However, as the proportion of resuscitation episodes that could successfully be observed was unknown we allocated a total period of 3 months for phase 1 observations in case it was required. In the absence of prior data we assumed resuscitation practices were appropriate on average on 50% (standard deviation $\pm 7.5\%$) of occasions. Further assuming an intra-class correlation coefficient (ICC) of 0.15, [7] a two-tailed test at the 5% significance level and 90% power, we estimated that a minimum of 22 health workers in each arm would need to be monitored with 4 observations made on each (i.e. at least 88 resuscitation events in both intervention

and control groups) to detect a 25% absolute change in our primary outcome measure (a 50% improvement) [8]. As these assumptions were based on limited data, particularly with regard to the frequency of our primary outcome and the value of the intra-class correlation coefficient we aimed to train at least 28 health workers in the first training and observe practices for these and for as many of the untrained providers as possible within the practice observation period.

Data Analysis

All observation checklist data were double entered using MS Access and verified prior to analysis using STATA v.9.2 (Stata Corp., Texas, USA). Two investigators and NRT instructors (ME and FW), blinded to the health workers' identity or training status, independently assigned a score to each resuscitation episode based on review of all of the information on the observation sheet and using a 5 point scale, where 5 represented perfect resuscitation (see Appendix S3). Scores were compared and individual cases where scores differed by >1 point were discussed by the two investigators with a revised, agreed final score applied. For cases where scores differed by ≤ 1 point the average of the two scores was considered the final score.

Observations were linked by the unique health worker study code and all analyses accounted for non-independence. Our analysis took into consideration the clustered nature of data in that health workers cared for more than one neonate. We used a cluster adjusted chi-square test to compare the proportions of appropriate initial resuscitation steps between the intervention and control groups. For the frequency of inappropriate practices and to compare the mean score for resuscitation performance we used a cluster adjusted two sample *t* test. We report risk ratios (RR) and 95% confidence intervals (CIs) (also adjusted for clustering) for the primary outcome. Confounding was explored for the categorical variables sex, years of experience (categorised as \geq median or <median) and place of work (labour ward or theatre) by calculating stratified, cluster adjusted risk ratios. After adjusting for these potential confounders there was no clinical or statistically significant variation in the main outcome of interest.

Ethics

The study was conducted with the permission of the hospital management to whom we explained the implications, purpose and voluntary nature of participation. Similar information was made available in written form to all labour ward and theatre staff and written informed consent was obtained from all health workers prior to their practice being observed. Information on the nature and purpose of the study and the need for the presence of an observer was also given to mothers admitted to the hospital for delivery. Mothers were given the opportunity to decline the presence of a resuscitation observer. As this hospital serves a national population of almost 3 million people we did not attempt to gain 'community consent' outside the hospital. Ethical approval for the conduct of the study was obtained from the Kenya Medical Research Institute / National Ethics Committee.

Results

Although our intention was to randomise staff, stratified by place of work (labour ward or theatre), to early or late training this proved to be impossible for the most part as a large number of potentially eligible staff did not meet our inclusion criteria because of expected absences of >2 weeks in the 3 months observation period for leave, scheduled off-duty periods or attendance at training seminars (figure 1). The final allocation of participants and process of observation is summarised in figure 1. Most of the

providers were females (trained; females 89.3 % (25/28), males 10.7 % (3/28), untrained; females 78.2% (43/55), males 21.8 % (12/55). There were no significant differences in the ages (median age (interquartile range, IQR); trained, 36(27–47), untrained, 35(27–51) and years of experience between the groups with the majority of health workers being relatively junior (median years worked (IQR), trained 1(1–20), untrained 1(1–20). Two hundred and twelve resuscitation episodes were observed for 83 providers in phase 1 while 50 were from 34 providers in phase 2. Ninety seven of the phase 1 practices were from 28 trained providers while 115 were from 55 untrained providers. Thirty five of the phase 2 practices were from 23 trained providers while 15 were from 11 remaining untrained providers. The profile of study patients and nursery admissions and deaths is summarised in table 1.

For our primary outcome in phase 1, we observed a significantly higher proportion of perfect initial resuscitation steps (24%) among trained providers compared to the control group (10%) (Risk ratio [RR] 2.27, 95% CI 1.23–4.22; $p=0.009$, adjusted for clustering) (Table 2). Similarly, the proportion of adequate initial resuscitation steps was higher among trained (66%) providers as compared to the control group (27%) (RR 2.45, 95% CI 1.75–3.42; $p<0.001$, adjusted for clustering). Analyses taking account of a possible confounding effect of the baseline imbalance in gender did not alter the observed effect of training; adequate resuscitation, RR 2.34, 95% CI 1.67–3.27, $p<0.001$, adjusted for sex and clustering). Results from analyses based on pooled data from both phase 1 and 2 periods were similar (Table 2). Risk ratios calculated for individual time periods each representing one third of the follow-up time in phase 1 did not demonstrate any converging trend (data not shown), arguing against a significant effect of contamination, although clearly there was limited power to detect anything but a major effect.

Similarly comparisons of trained and untrained providers for phase 1 and phase 1 and 2 combined showed significantly fewer inappropriate and potentially harmful practices (summarised in Appendix S2) per resuscitation in the trained group (Phase 1: Mean difference 0.40 (trained 0.53 vs untrained 0.92), 95% CI 0.13–0.66; $p=0.0038$) (Table 3). A total of 256 (98.0%) resuscitation episodes were documented sufficiently well to permit scoring. Phase 1 group comparison showed significantly higher average resuscitation scores in the trained group as compared to the control group (Mean score: trained 2.50, 95% CI 2.25–2.74; untrained 1.95, 1.74–2.16, $p=0.0008$). This effect was also apparent using pooled data from Phase 1 and 2 (Table 3). In consecutive observations in the period after late training the proportion of resuscitation episodes with adequate initiation of resuscitation was 70% (95% CI 51.4%–88.7%).

Group comparison for the overall mortality in all the resuscitation episodes showed no statistically significant differences between the groups (Trained 0.28 (18/65), 95% CI 0.17–0.40; control 0.25 (9/25), 0.12–0.42, $p=0.77$). Additionally, no significant differences were seen in birth asphyxia admission and fatality rates before and after training (Table 1). For birth asphyxia pre-intervention admission rates to the newborn unit among infants weighing 2000–4000g were 13.1% of all births, 95% CI 12.4%–13.8% while post-intervention they were 11.7%, 11.0%–12.4%. Fatality rate amongst infants weighing >2000g admitted to the newborn unit with asphyxia was 6.4%, (5.1%–8.0%) in the pre-intervention period and 6.6% (5.1%–8.3%) in the period following late training.

Discussion

We attempted to undertake a cluster-randomised trial to study the effect of a simple one day newborn resuscitation training on health worker practices. However, our criteria for randomisation,

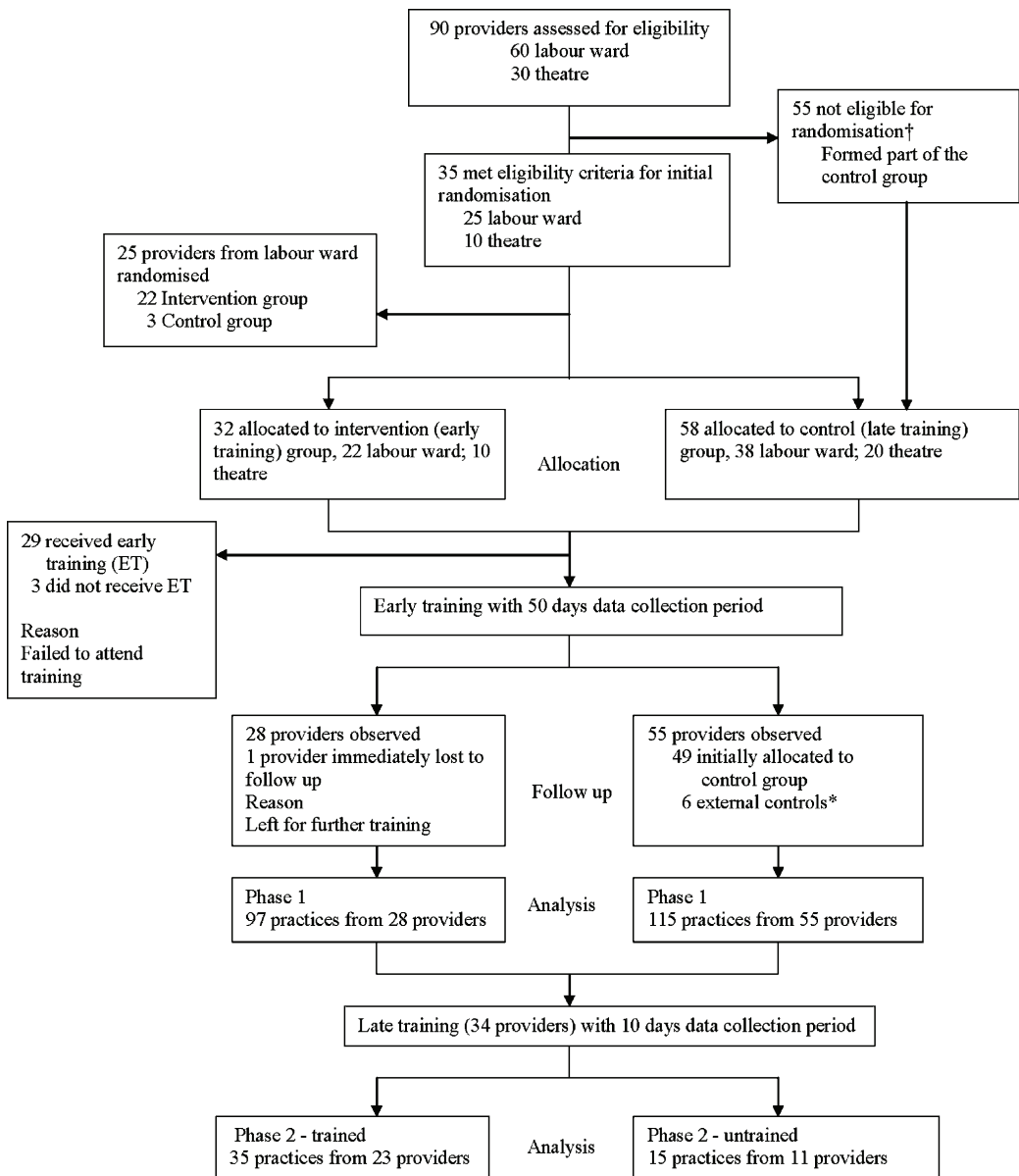


Figure 1. Trial profile

† Personal work plans for the 3 months post-randomisation included leave of >2 weeks

* Providers not initially allocated but with observed practices.

Figure 1. Trial profile
doi:10.1371/journal.pone.0001599.g001

Table 1. Profile of study patients

	Pre-intervention phase		Post-intervention phase	
	Period 1	Period 2	Period 3	Period 4
Number of deliveries	4367	4302	4205	4084
Stillbirths				
Fresh	67	80	69	54
Macerated	64	66	60	62
Neonatal deaths†	7	9	5	7
Birthweights				
<2000 g	213	223	194	211
2000–2499 g	362	339	286	312
2500–4000 g	3668	3663	3667	3629
>4000 g	102	72	70	82
Illness specific nursery admissions and deaths				
Birth asphyxia				
<2000 g	66(13)	35(8)	21(7)	19(3)
2000–2499 g	75(23)	92(10)	54(1)	51(5)
2500–4000 g	474(23)	495(17)	441(20)	426(38)
>4000 g	23(0)	17(0)	16(1)	9(0)
Prematurity	152(37)	165(34)	137(42)	197(39)
RDS (Term)	44(0)	48(6)	37(7)	80(12)
RDS (Preterm)	12(5)	11(3)	12(5)	8(2)
Neonatal sepsis	40(0)	46(1)	19(0)	35(1)
Jaundice	25(0)	33(2)	26(1)	11(0)
MAS	0(0)	0(0)	2(0)	6(0)
Congenital abnormality	14(1)	16(1)	17(1)	21(0)
Neonatal mortality rate* (95% CI)	25.0 (20.5–30.0)	21.2 (17.1–25.9)	21.4 (17.3–26.2)	26.2 (21.5–31.6)

†Deaths during resuscitation;

*In-hospital rate per 1000 live births

RDS: respiratory distress syndrome; MAS: meconium aspiration syndrome

Deaths are given in parentheses

doi:10.1371/journal.pone.0001599.t001

aiming to ensure health workers were present to be observed in a defined period, resulted in few staff being eligible. We cannot therefore discount the possibility of bias in group allocation although we feel this is unlikely. The training intervention significantly improved the performance of initial resuscitation steps, with 66% initial practices being adequate in the intervention group compared with 27% in the control group. In addition, there were significant reductions in the frequency of inappropriate and potentially harmful practices and improvements in overall resuscitation scores. There was no obvious effect of training on mortality of babies resuscitated, no obvious decline in asphyxia admission rates and no overall decline in newborn mortality in the hospital as the number of trained providers increased. However, this study was neither specifically designed nor powered with mortality as the primary outcome and our mortality results are best used to inform the design of future studies. In addition, appropriate initial resuscitation is clearly only the first stage in a continuum of effective care, not addressed by this intervention, that is likely to be required to prevent many adverse outcomes from severe asphyxia.

We are not aware of any previous randomised controlled studies examining the effect of resuscitation training on provider practices in a true clinical setting. The majority of studies on newborn

Table 2. Group comparison for appropriate initial resuscitation steps (all analyses are cluster adjusted)

	Mean	Risk ratio (95% CI)	p-value
Phase 1			
Perfect resuscitation	23.7%/10.4%	2.27 (1.23–4.22)	0.009
Adequate resuscitation	66.0%/27.0%	2.45 (1.75–3.42)	<0.001
Phase 2			
Perfect resuscitation	40.0%/13.3%	3.00 (0.79–11.42)	0.064
Adequate resuscitation	74.3%/60.0%	1.24 (0.71–2.15)	0.312
Phase 1 and 2			
Perfect resuscitation	28.0%/10.8%	2.60 (1.53–4.43)	<0.001
Adequate resuscitation	68.1%/30.8%	2.22 (1.64–2.99)	<0.001

doi:10.1371/journal.pone.0001599.t002

resuscitation have focussed on less direct outcomes such as participants' knowledge and skills [5,9,10]. Such surrogate outcomes may not necessarily reflect practice changes, a more useful and direct way of measuring the effectiveness of resuscitation training programmes [5]. Although our primary study outcome was only able to capture the initial steps in effective practice we believe it does indicate an important behaviour change effect, especially if considered together with the reduction in unnecessary / potentially harmful practices and an improvement in overall resuscitation scores.

Table 3. Mean number of inappropriate/harmful practices and resuscitation scores per episode (all analyses are cluster adjusted)

	N	Clusters	Mean (95% CI)	p-value
a) Inappropriate and dangerous practices				
Phase 1				
Intra-cluster correlation = 0.20				
Untrained = 0	115	55	0.92 (0.75–1.10)	
Trained = 1	97	28	0.53 (0.32–0.73)	
Difference (0–1)	212	83	0.39 (0.13–0.66)	0.0038
Phase 1 and 2				
Intra-cluster correlation = 0.19				
Untrained = 0	130	61	0.87 (0.72–1.02)	
Trained = 1	132	51	0.45 (0.29–0.61)	
Difference (0–1)	262	112	0.42 (0.21–0.64)	0.0002
b) Mean resuscitation scores				
Phase 1				
Intra-cluster correlation = 0.12				
Untrained = 0	112	54	1.95 (1.74–2.16)	
Trained = 1	94	28	2.50 (2.25–2.74)	
Difference (0–1)	206	82	–0.55 (–0.86, –0.23)	0.0008
Phase 1 and 2				
Intra-cluster correlation = 0.12				
Untrained = 0	127	60	1.83 (1.61–2.04)	
Trained = 1	129	51	2.40 (2.18–2.61)	
Difference (0–1)	256	111	–0.57 (–0.87, –0.27)	0.0003

doi:10.1371/journal.pone.0001599.t003

Previous studies and our control group data demonstrate that both resuscitation skills and knowledge are poor despite frequent exposure to situations in which both are needed [3,11]. Internationally, there is now considerable consensus on how newborn resuscitation should be provided [12] and it is believed that in 95% cases when it is required resuscitation should be possible with only a minimum of equipment and without access to intensive care skills or facilities [4,13]. Recent research findings have strengthened this opinion demonstrating that suction in the presence of meconium and the use of oxygen are in most newborns unnecessary [14,15,16,17]. These findings have relevance to our study as the failure to provide suction to a non-breathing baby born through meconium as the first step was a major reason for failing to achieve a 'perfect' classification in our primary outcome. If, as seems likely, there is little value of suction in these babies then a substantial clinical impact from our intervention, 66% of adequate appropriate practices in trained providers, might be a more reasonable interpretation than the modest impact suggested by only 25% of initial practices in trained providers being perfect.

Our data add to a body of knowledge suggesting some improvement in clinical outcomes [10,18] or in acquisition of knowledge and skills of providers following resuscitation training [9]. In a systematic review on the effectiveness of all types of life support courses all the three mortality and morbidity studies indicated a positive impact, with an overall odds ratio of 0.28 (95% CI 0.22–0.37). However, no net increase in scores in 5/8 studies of retention of knowledge and in 8/9 studies of skills retention were apparent, although all the studies assessing behavioral outcomes were reported to be methodologically weak [5].

Similarly, our study has limitations. Attempts to randomise health workers had limited success. We cannot exclude the possibility of cross-group contamination, although this would tend to reduce the apparent effect of the intervention. In contrast it is likely that the difficulty in maintaining observer blinding could bias the results in favour of an intervention effect. If the observers, even unintentionally, were more likely to view the practices of a provider they came to know was trained as correct this would bias our results despite our efforts in training to limit this effect. We also only observed practitioners for a short period after training and are unable to provide any information on the duration of the training effect. In the few studies assessing the duration of effect a rapid and linear decay in cardio-pulmonary (CPR) skills -from as early as two weeks after training, with skills deteriorating to pre-training levels by one year, have been reported [11,19,20,21].

For low-income countries Life Support Courses are associated with relatively high direct and opportunity costs (learners'/instructors' time, equipment purchase, etc). While there is increasing pressure to implement such courses it is important that their true effects on actual health worker performance and ideally morbidity and mortality are established. Such studies need to be based in typical, low-income settings where supervision and opportunities for continuous learning or ongoing mentorship and resources for post-resuscitation care may be limited. In addition, they should perhaps consider a range of possible training delivery mechanisms, be

embedded in local health systems to promote sustainability, assess impact over the long term and consider costs and cost effectiveness to optimise appropriate health policy decisions. Clearly such studies will require appropriate levels of funding.

In conclusion, our findings suggest that implementation of a simple one day newborn resuscitation training can be followed by significant, short-term improvement in health workers' practices. To ensure a high proportion of all resuscitation episodes are appropriately managed clearly a large majority of providers must be trained. Evidence on effects on long term performance or clinical outcomes, however, remain inconclusive and can only be established by larger trials. The availability, accessibility and correct functioning of basic resuscitation equipment is still a missing essential pre-requisite for the success of training and resuscitation itself in many settings [2].

Supporting Information

Checklist S1 CONSORT Checklist

Found at: doi:10.1371/journal.pone.0001599.s001 (0.25 MB PDF)

Protocol S1 Trial Protocol

Found at: doi:10.1371/journal.pone.0001599.s002 (0.06 MB PDF)

Appendix S1 Levels of appropriate initial resuscitation steps

Found at: doi:10.1371/journal.pone.0001599.s003 (0.04 MB RTF)

Appendix S2 Inappropriate and harmful practices

Found at: doi:10.1371/journal.pone.0001599.s004 (0.04 MB RTF)

Appendix S3 Scoring instrument for the assessment of resuscitation practices

Found at: doi:10.1371/journal.pone.0001599.s005 (0.04 MB RTF)

Acknowledgments

This work is published with the permission of the Director of KEMRI. We thank all the nurses, resuscitation observers and NLS instructors who contributed to the study. We would also like to thank the Pumwani Hospital administration for their support in the conduct of this study and the mothers who agreed to allow observers to be present at the delivery of their babies.

Author Contributions

Conceived and designed the experiments: ME. Analyzed the data: GF ME NO FW. Contributed reagents/materials/analysis tools: GF ME NO FG AW. Wrote the paper: NO. Other: Contributed to the study development and implementation of the NLS course: NO FW FG AW ME. NLS instructor in the study: FG AW ME. Supervised data collection and controlled the quality of the study: NO ME. Contributed to the analysis and interpretation of the results: FW ME. Reviewed and approved the final version of the manuscript: FW FG GF AW ME. Contributed to the design of the study and the statistical analysis and interpretation of the results: GF. Designed and obtained funding for the study: ME NO.

References

1. WHO (2005) The world health report: 2005: make every mother and child count.
2. English M, Esamai F, Wasunna A, Were F, Ogutu B, et al. (2004) Delivery of paediatric care at the first-referral level in Kenya. *Lancet* 364: 1622–1629.
3. Kamnir S (1997) Neonatal resuscitation and newborn outcomes in rural Kenya. *J Trop Pediatr* 43: 170–173.
4. WHO (1997) Basic newborn resuscitation: a practical guide. Geneva: World Health Organisation.
5. Jabbour M, Osmond M, Klassen T (1996) Life support courses: are they effective? *Ann Emerg Med* 28: 690–698.
6. Baskett P, Nolan J, Handley A, Soar J, Biarent D, et al. (2005) European resuscitation council guidelines for resuscitation 2005. Section 9. Principles of training in resuscitation. *Resuscitation* 67S1: S181–S189.

7. Rowe A, Lama M, Onikpo F, Deming M (2002) Design effects and intraclass correlation coefficients from a health facility cluster survey in Benin. *Int J Qual Health Care* 14: 521–523.
8. Hayes R, Bennett S (1999) Simple sample size calculation for cluster-randomized trials. *Int J Epidemiology* 28: 319–326.
9. Durojaiye L, O'meara M (2002) Improvement in resuscitation knowledge after a one-day paediatric life-support course. *J Paediatr Child Health* 38: 241–245.
10. Quan L, Shugerman R, Kunkel N, Brownlee C (2001) Evaluation of resuscitation skills in new residents before and after pediatric advanced life support course. *Pediatrics* 108: E110.
11. Carbine D, Finer N, Knodel E, Rich W (2000) Video recording as a means of evaluating neonatal resuscitation performance. *Pediatrics* 106: 654–658.
12. International Liaison Committee on Resuscitation (2005) 2005 International Consensus on Cardiopulmonary Resuscitation and Emergency Cardiovascular Care Science with Treatment Recommendations. Part 7: Neonatal resuscitation. *Resuscitation* 67.
13. Opiyo N, English M (2006) Newborn resuscitation: defining best practice for low-income settings. *Trans R Soc Trop Med Hyg* 100: 899–908.
14. Halliday H, Sweet D (2001) Endotracheal intubation at birth for preventing morbidity and mortality in vigorous, meconium-stained infants born at term (Review). *The Cochrane Database of Syst Rev* 2: Issue 1.
15. Tan A, Schulze A, O'Donnell C, Davis P (2004) Air versus oxygen for resuscitation of infants at birth (Review). *The Cochrane Database of Syst Rev* Issue 3.
16. Vain N, Szyld E, Prudent L, Wiswell T, Aguilar A, et al. (2004) Oropharyngeal and nasopharyngeal suctioning of meconium-stained neonates before delivery of their shoulders: multicentre, randomised controlled trial. *Lancet* 364: 597–602.
17. Wiswell T, Gannon C, Jacob J, Goldsmith L, Szyld E, et al. (2000) Delivery room management of the apparently vigorous meconium-stained neonate: results of the multicentre, international collaborative trial. *Pediatrics* 105: 1–7.
18. O'Hare B, Nakakeeto M, Southall D (2006) A pilot study to determine if nurses trained in basic neonatal resuscitation would impact the outcome of neonates delivered in Kampala, Uganda. *J Trop Pediatr* 52: 376–379.
19. Berden H, Willems F, Hendrick J, Pijls M, Knape J (1983) How frequently should basic cardiopulmonary resuscitation training be repeated to maintain adequate skills? *BMJ* 306: 1576–1577.
20. Jewkes F, Phillips B (2003) Resuscitation training of paediatricians. *Arch Dis Child* 88: 118–121.
21. McKenna S, Glendon A (1985) Occupational first aid training: decay in cardiopulmonary resuscitation (CPR) skills. *Journal of occupational psychology* 58: 109–117.

Appendix S3. Scoring instrument for the assessment of resuscitation practices

Score	Interpretation
5	Resuscitation entirely appropriate - no inappropriate practices
4	Resuscitation good- no inappropriate practices, minor, clinically insignificant deviations from recommended sequence
3	Resuscitation adequate and not dangerous - inappropriate practices do not interfere with care or threaten outcome, deviations from recommended sequences unlikely to be of significance
2	Resuscitation poor- inappropriate practices or deviations from recommended sequences could have some impact on outcome (significant delay in establishing adequate ventilation, potential for mild adverse effect)
1	Resuscitation very poor and potentially dangerous- inappropriate practices or deviations from recommended sequences significantly delay effective care in a very sick baby or can cause possibly serious adverse consequences

In-service training for health professionals to improve care of the seriously ill newborn or child in low and middle-income countries (Review) (Review)

Opiyo N, English M



This is a reprint of a Cochrane review, prepared and maintained by The Cochrane Collaboration and published in *The Cochrane Library* 2010, Issue 5

<http://www.thecochranelibrary.com>



In-service training for health professionals to improve care of the seriously ill newborn or child in low and middle-income countries
(Review) (Review)
Copyright © 2010 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

TABLE OF CONTENTS

HEADER	1
ABSTRACT	1
PLAIN LANGUAGE SUMMARY	2
SUMMARY OF FINDINGS FOR THE MAIN COMPARISON	2
BACKGROUND	5
OBJECTIVES	7
METHODS	7
RESULTS	9
Figure 1.	10
Figure 2.	11
Figure 3.	11
DISCUSSION	11
AUTHORS' CONCLUSIONS	12
ACKNOWLEDGEMENTS	13
REFERENCES	13
CHARACTERISTICS OF STUDIES	15
DATA AND ANALYSES	18
Analysis 1.1. Comparison 1 Opiyo 2008, Outcome 1 Mortality.	18
Analysis 2.1. Comparison 2 Senarath 2007, Outcome 1 Practice of preparedness of resuscitation.	19
Analysis 2.2. Comparison 2 Senarath 2007, Outcome 2 Preparedness for resuscitation - adjusted for clustering.	19
APPENDICES	19
WHAT'S NEW	32
HISTORY	32
CONTRIBUTIONS OF AUTHORS	33
DECLARATIONS OF INTEREST	33
SOURCES OF SUPPORT	33
INDEX TERMS	33

[Intervention Review]

In-service training for health professionals to improve care of the seriously ill newborn or child in low and middle-income countries (Review)

Newton Opiyo¹, Mike English¹

¹Child and Newborn Health Group, Kenya Medical Research Institute/Wellcome Trust Research Programme, Nairobi, Kenya

Contact address: Newton Opiyo, Child and Newborn Health Group, Kenya Medical Research Institute/Wellcome Trust Research Programme, PO Box 43640, Nairobi, 00100 GPO, Kenya. nopiyo@nairobi.kemri-wellcome.org.

Editorial group: Cochrane Effective Practice and Organisation of Care Group.

Publication status and date: Edited (no change to conclusions), published in Issue 5, 2010.

Review content assessed as up-to-date: 21 May 2009.

Citation: Opiyo N, English M. In-service training for health professionals to improve care of the seriously ill newborn or child in low and middle-income countries (Review). *Cochrane Database of Systematic Reviews* 2010, Issue 4. Art. No.: CD007071. DOI: 10.1002/14651858.CD007071.pub2.

Copyright © 2010 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

ABSTRACT

Background

A variety of emergency care training courses based on developed country models are being promoted as a strategy to improve the quality of care of the seriously ill newborn or child in developing countries. Clear evidence of their effectiveness is lacking.

Objectives

To investigate the effectiveness of in-service training of health professionals on their management and care of the seriously ill newborn or child in low and middle-income settings.

Search strategy

We searched The Cochrane Register of Controlled Trials (CENTRAL), the Specialised Register of the Cochrane EPOC group (both up to May 2009), MEDLINE (1950 to May 2009), EMBASE (1980 to May 2009), CINAHL (1982 to March 2008), ERIC / LILACS / WHOLIS (all up to October 2008), and ISI Science Citation Index Expanded and ISI Social Sciences Citation Index (both from 1975 to March 2009). We checked references of retrieved articles and reviews and contacted authors to identify additional studies.

Selection criteria

Randomised controlled trials (RCTs), cluster-randomised trials (CRTs), controlled clinical trials (CCTs), controlled before-after studies (CBAs) and interrupted time series studies (ITSs) that reported objectively measured professional practice, patient outcomes, health resource /services utilization, or training costs in healthcare settings (not restricted to studies in low-income settings).

Data collection and analysis

We independently selected studies for inclusion, abstracted data using a standardised form, and assessed study quality. Meta-analysis was not appropriate. Study results were summarised and appraised.

In-service training for health professionals to improve care of the seriously ill newborn or child in low and middle-income countries (Review) (Review)

Copyright © 2010 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Main results

Two studies of varied designs were included. In one RCT of moderate quality, Newborn Resuscitation Training (NRT) was associated with a significant improvement in performance of adequate initial resuscitation steps (risk ratio 2.45, 95% confidence interval (CI) 1.75 to 3.42, $P < 0.001$, adjusted for clustering) and a reduction in the frequency of inappropriate and potentially harmful practices (mean difference 0.40, 95% CI 0.13 to 0.66, $P = 0.004$). In the second RCT, available limited data suggested that there was improvement in assessment of breathing and newborn care practices in the delivery room following implementation of Essential Newborn Care (ENC) training.

Authors' conclusions

There is limited evidence that in-service neonatal emergency care courses improve health-workers' practices when caring for a seriously ill newborn although there is some evidence of benefit. Rigorous trials evaluating the impact of refresher emergency care training on long-term professional practices are needed. To optimise appropriate policy decisions, studies should aim to collect data on resource use and costs of training implementation.

PLAIN LANGUAGE SUMMARY

Effectiveness of in-service training in the care of the seriously ill newborn or child

In developing countries, most deaths in very ill babies and children who seek care in healthcare facilities happen within 48 hours of being seen. Currently, a number of emergency care courses, adapted from developed countries are being promoted in developing countries as a means to improving the quality of care provided to seriously ill newborns or children. Whether these courses result in improvement in health workers' ability to provide appropriate care remains unclear.

Although the results from the two included studies showed that emergency care training could be followed by improvement in health workers' practices, because of the small number of studies, differences in the training courses, and weaknesses in the study methods, it is not possible to conclude that in general such in-service training improves health worker practices when they are faced with a seriously ill child. Further well-conducted studies are therefore needed to provide reliable evidence on what such courses achieve. To guide decisions regarding which interventions to invest in, such studies should also collect data on resources used and costs of training interventions.

SUMMARY OF FINDINGS FOR THE MAIN COMPARISON [Explanation]

In-service neonatal emergency care training versus standard care for healthcare professionals							
Patient or population: Healthcare professionals Settings: Hospital-based settings Intervention: In-service neonatal emergency care training Comparison: Usual or standard care							
Outcomes	Illustrative comparative risks* (95% CI)			Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk					
	Usual or standard care	In-service neonatal emergency care training					
Proportion of appropriate initial resuscitation steps Direct observations Follow-up: 50 days				RR 2.45 (1.75 to 3.42)	83 [1 study]	+++0 moderate	Overall quality: high risk of bias
	27 per 1000	67 per 100 (48 to 93)					
	Medium risk population						
Frequency of inappropriate practices Follow-up: mean 50 days	The mean frequency of inappropriate practices in the control group was 0.92		The mean frequency of inappropriate practices in the control group was 0.40 higher (0.13 to 0.66 higher)	83 [1 study]	+++0 moderate		
Mortality Direct observations Follow-up: mean 40 days				RR 0.77 (0.40 to 1.48)	90 (1 study)	+++0 moderate	Trial not sufficiently powered to detect a mortality effect
	36 per 100	28 per 100 (14 to 53)					
	Medium risk population						

<p>*The basis for the assumed risk (e.g. the median control group risk across studies) is provided in footnotes. The corresponding risk (and its 95% confidence interval) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI).</p> <p>CI: Confidence interval; RR: Risk Ratio</p>	
<p>GRADE Working Group grades of evidence</p> <p>High quality: Further research is very unlikely to change our confidence in the estimate of effect.</p> <p>Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.</p> <p>Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.</p> <p>Very low quality: We are very uncertain about the estimate.</p>	

BACKGROUND

In developing countries most deaths among seriously ill children who come into contact with referral level health services occur within 48 hours of being seen (Berkley 2005). It is possible that good quality immediate and effective care provided by health professionals could reduce these deaths (Nolan 2001). Provision of appropriate care is, however, dependent on the presence of adequately skilled health personnel at the point of delivery (WHO 2005). To improve health workers' capacity to provide effective care for seriously ill children and newborns in low-income countries, a number of in-service training courses, mainly based on developed countries' models, are proposed.

These courses include: (1) neonatal life support courses (e.g. Newborn Life Support (NLS), Neonatal Resuscitation Program (NRP)), (2) paediatric life support courses (e.g. Paediatric Advanced Life Support (PALS), Paediatric Life Support (PLS)), (3) life support / emergency care elements within the Integrated Management of Pregnancy and Childbirth (e.g. Essential Newborn Care (ENC)) and (4) components of other in-service child health training courses that deal with care of serious illness (e.g. Emergency Triage, Assessment and Treatment (ETAT), Control of Diarrheal Diseases (CDD) and Acute Respiratory Infections (ARI)

case management programs and the training components of the Integrated Management of Childhood Illness (IMCI) strategy).

Although such formalised educational programs vary in origin, scope and target audience, they are typically aimed at in-service rather than pre-service training, and are short and intensive with a structured approach to the presentation of their clinical subject (Table 1). The one-day NRP course was first taught in 1987 in the USA while the one-day NLS course was initiated in the UK in 2001 (Raupp 2007). PALS, a two-day course, was piloted in USA in 1988. Advanced Paediatric Life Support (APLS), a three-day course, was developed and piloted in the UK in 1992. Two other courses - the one-day PLS course and Prehospital PLS - have been designed to complement the APLS (Jewkes 2003). The World Health Organization (WHO) has recently added to this list the 3½-day ETAT course based upon and validated against the APLS course in Malawi (Gove 1999; Molyneux 2006). This course is aimed specifically at lower income settings and is intended to improve prompt identification and institution of life saving emergency treatment for very ill children. These life support courses emphasize early recognition of neonatal/paediatric emergencies and prevention of cardio-respiratory arrest (and mortality) through resuscitation.

Table 1. Summary of neonatal and paediatric emergency care courses†

Course	Subject	Duration (days)	Target audience
Neonatal Life Support (NLS)	Neonatal resuscitation	1	Midwives Paediatricians General Practitioners
Neonatal Resuscitation Program (NRP)	Neonatal resuscitation	1	Midwives Paediatricians General practitioners
Paediatric Life Support (PLS)	Basic Life Support (BLS) and Advanced Life Support (ALS) for children Recognition of paediatric emergencies	1	Nurses and doctors involved in paediatric care
Paediatric Advanced Life Support (PALS)	BLS and ALS for children Recognition of paediatric emergencies Some neonatal life support	2	Nurses and doctors involved in paediatric care
Prehospital Paediatric Life Support (PHPLS)	Prehospital paediatric emergency care	2+	General practitioners, paramedics, some nurses, emergency medicine staff

Table 1. Summary of neonatal and paediatric emergency care courses† (Continued)

Advanced Paediatric Life Support (APLS)	BLS and ALS for children Paediatric emergencies, including serious illness and major trauma, some neonatal life support	3	Paediatricians, emergency medicine doctors, some anaesthetists, senior paediatric nurses
Emergency Triage Assessment and Treatment (ETAT)	Very ill children presenting to hospital	3.5	Doctors, nurses, paramedics
Essential Newborn Care (ENC) course	Aspects of newborn care (including neonatal resuscitation) in the Integrated Management of Pregnancy and Childbirth (IMPAC)	5	Nurses, midwives, doctors
Integrated Management of Childhood Illness (IMCI)	Ill children and neonates including emergency care or identification and referral of the seriously ill	11	Nurses, midwives, doctors

†Tulloch 1999, Jewkes 2003, Mello 2003, Irimu 2008

The more general CDD and ARI programs were developed by the WHO in 1980, in recognition of the high childhood mortality due to diarrhea/dehydration and pneumonia for the very ill neonate or child and focus on case management training rather than life-support (Forsberg 2007; Pio 2003). While these courses concentrate predominantly on community or outpatient based management, where there is good evidence for their success (Sazawal 2001), they also include guidance on management of very severe illness. These disease-specific training approaches were incorporated into the broader package of the IMCI strategy. Here the particular focus for management of the very ill child is the decision to refer to hospital and provide pre-referral management. In addition to this, the WHO has developed a specific five-day course on hospital management of severe malnutrition (WHO 2002).

In-service training, however, costs both time and money: for example, the cost of the 2-day European Paediatric Life Support (EPLS) course is estimated to be about USD 190 per trainee in Kenya (Personal communication with ME, 2009). These costs include allowances for the trainers (e.g. travel refunds), course materials (e.g. course manuals, consumable teaching aides, etc) and hotel costs for the participants. Apart from the sometimes high costs of providing courses (recovered in high income countries often with high course fees), attendance at such courses often means that important staff (instructors and participants) are absent from their normal duties with potential disruption to patient care and for

some a loss of personal income (Jabbour 1996). Despite their cost, however, emergency care courses are a thriving enterprise in many high income countries, as reflected in their ever increasing number and variety (Jewkes 2003). In the hope that they might improve the quality of care in many low- and middle-income countries, considerable global efforts and investments have gone into their further development, refinement and adaptation to meet individual country needs (Baskett 2005). Yet despite these investments and the faith placed in them by many organizations and institutions, clear evidence of the effectiveness of these courses in improving health workers ability to manage seriously ill children or neonates appears lacking.

Two Cochrane reviews from the Injuries Group have examined the effectiveness of Advanced Trauma Life Support course (ATLS), for ambulatory crews (Sethi 2001) and hospitals (Shakiba 2003), respectively, in reducing mortality and morbidity for victims of trauma of any age. The pre-hospital review (Sethi 2001) identified one small randomised controlled trial (n = 16) while the hospital review (Shakiba 2003) did not find any relevant randomised controlled trials. The Cochrane review authors concluded that there is no clear evidence that ATLS training impacts on the outcome of victims of trauma.

The effectiveness of in-service training of health professionals depends on changes in health worker practices which, plausibly,

should precede any impact on mortality or morbidity. This review investigated if there are systematic differences in health workers' professional practices (i.e. more appropriate management or referral of seriously ill children/newborns or both), patient outcomes (mortality and morbidity), or health resource use (e.g. drug use, laboratory tests) and services utilization (e.g. length of hospitalization, return visits) after in-service training in emergency care or care for the seriously ill newborn or child. Information regarding the effectiveness of such in-service training courses is required to enable low-income countries to prioritise the health interventions they invest in.

OBJECTIVES

To investigate the effectiveness of in-service training of health professionals on their management and care of seriously ill neonates or children in low-income settings.

We considered the comparisons listed below.

1. Interventions in which seriously ill neonates have been cared for by qualified health professionals who have undergone neonatal emergency care training compared to those receiving usual or standard care.
2. Interventions in which seriously ill children have been cared for by qualified health professionals who have undergone paediatric emergency care training compared to those receiving usual or standard care.
3. Interventions in which seriously ill neonates or children have been cared for by qualified health professionals who have undergone any other in-service child health training that deals with care of severe illness (e.g. CDD, ARI, ETAT, etc) compared to those receiving usual or standard care.

METHODS

Criteria for considering studies for this review

Types of studies

All randomised controlled trials (RCTs), cluster-randomised trials (CRTs), controlled clinical trials (CCTs), controlled before-after studies (CBAs) and interrupted time series studies (ITSs) that have evaluated the effects of in-service training on at least one of the outcomes listed below. We did not consider before and after studies that had no parallel control groups.

Types of participants

All qualified healthcare professionals, including, doctors (general practitioners and specialists), nurses, pharmacists and dietitians/nutritionists, in outpatient or hospital-based settings, responsible for the management and care of seriously ill neonates or children. We excluded non-qualified healthcare providers (e.g. medical students/trainees, medical interns, community health workers). Studies were not excluded based on the setting (low or high income).

Types of interventions

We considered implementation studies of the following in-service training courses aimed at changing provider behaviour in the care of the seriously ill newborn or child:

- 1) Neonatal life support courses e.g. NLS, NRP, and others.
- 2) Paediatric life support courses e.g. PALS, PLS, and others.
- 3) Life support elements within the Integrated Management of Pregnancy and Childbirth e.g. ENC.
- 4) Other in-service newborn and child health training courses aimed at the recognition and management of the seriously ill child e.g. ETAT, CDD, ARI, malaria case management or the training components of the IMCI strategy.

We excluded studies of complex interventions in which training is combined with and impossible to separate from additional health system improvements (for example improved drug/equipment/staff supply/health facility reorganisation).

Types of outcome measures

Primary outcomes

We included studies only if they reported at least one of the following objectively measured professional (in practice) performance outcomes.

- Adherence to treatment guidelines
- Prescribing practices
- Clinical assessment and diagnosis
- Recognition of and management or referral of the seriously ill newborn/child

Secondary outcomes

Where reported, we also considered the following outcomes

- Health resource utilization, for example, use of drugs, laboratory tests, etc.
- Health services utilization, for example, length of hospital stay.
- 'Other markers' of clinical performance, for example, simulated health worker performance (in practice)
- Cost of training, for example, costs of purchasing training materials.

We excluded studies that only reported 'other markers' of performance (for example, simulations/skill testing that is done outside of the practice setting (in the classroom) that are tests of skill, such as practicing/demonstrating resuscitation techniques using a dummy). However, we considered for inclusion simulations of emergency care in the practice setting that were designed to reflect real practice.

Search methods for identification of studies

See: Cochrane Effective Practice and Organisation of Care Group methods used in reviews.

We selected studies according to the Cochrane Handbook for Systematic Reviews of Interventions (Higgins 2006) and the Effective Practice and Organisation of Care Group (EPOC) methods used in reviews (EPOC 2007).

Electronic searches

To identify potential studies for inclusion, we searched the following electronic databases.

- a) The Cochrane Central Register of Controlled Trials (CENTRAL)/EPOC register (up to May 2009) (Appendix 1).
- b) MEDLINE (1950 to May 2009), EMBASE (1980 to May 2009), CINAHL (1982 to March 2008), LILACS (up to October 2008), ERIC (up to October 2008) and WHOLIS (up to October 2008). A forward search for papers that cite included studies was conducted in the ISI Science Citation Index Expanded (SCI-Expanded) (1975 to March 2009) and ISI Social Sciences Citation Index (SSCI) (1975 to March 2009).

We developed search strategies for electronic databases using the methodological component of the EPOC search strategy combined with selected MeSH terms and free text terms. Appendix 2 shows the terms used in the MEDLINE search strategy. We modified this search strategy as appropriate for other databases (Appendix 3; Appendix 4; Appendix 5; Appendix 6; Appendix 7). No date or language restrictions were applied.

Searching other resources

- a) List of references from the Health Care Provider Performance (HCPP) Review (Rowe 2008, available from Alexander K. Rowe, e-mail: axr9@cdc.gov).
- b) Clinical trial registries for ongoing studies
- c) Reviewed reference lists of all papers and relevant reviews identified.
- d) Contacted authors of relevant articles regarding any further published or unpublished work.

Data collection and analysis

Selection of studies

The two review authors independently screened the titles and abstracts (where available) based on the pre-determined review criteria. We retrieved all full text copies of studies meeting the inclusion criteria for a detailed assessment by both authors. Disagreements were resolved through consensus following discussion between the authors.

Data extraction and management

Both authors independently extracted data from trial reports using a modified EPOC data collection tool from the EPOC group (EPOC 2007). We extracted data relating to the following items:

1. Participants (healthcare providers and patients): profession, number, age, years of experience, and number of episodes/practices performed by the included healthcare providers. We also collected data regarding the number and specific clinical problems of the included patients.
2. Intervention: type and duration of training and co-interventions (teachings aids, self-learning manuals, etc).
3. Type of targeted behaviour (general management of the problem).
4. Study designs and the key characteristics of the studies (setting, unit of allocation/analysis, length of post-intervention follow-up, and time lag between the intervention and post-intervention assessment).
5. Results grouped according to the primary and secondary outcomes specified above.

We resolved disagreements through discussion.

Assessment of risk of bias in included studies

Both review authors independently assessed the risk of bias of all included studies using the EPOC checklist for the assessment of methodological quality of studies (see EPOC module) and rated them into three classes: low (low risk of bias for all key domains), high (high risk of bias for one or more key domains) and unclear risk of bias (unclear risk of bias for one or more key domains) based on the assessment of the following: allocation sequence generation, allocation concealment, measurement of baseline outcomes, baseline characteristics of providers, blinding (participants, personnel and outcome assessors), completeness of follow-up (mainly related to follow-up of professionals), treatment of incomplete outcome data, and protection against selective outcome reporting and contamination. We resolved disagreements regarding the quality ratings through discussion between the two authors. Studies were not excluded based on their risk of bias. We assessed the overall quality of evidence using the Grading of Recommendations Assessment,

Development and Evaluation (GRADE) approach, see [Summary of findings for the main comparison](#).

Data synthesis

Identified eligible studies varied in design, focus (newborn resuscitation, essential newborn care), length (1 versus 4 days) and outcome measured. It was therefore not appropriate to combine the results of the studies. The results are therefore presented separately.

In one study ([Senarath 2007](#)), data analyses could have been improved: First, there was a 'unit of analysis error' - the unit of randomisation was the hospital but the unit of analysis was observed practices at delivery. Thus, the strength of the reported associations could have been over-estimated ('spuriously low P values') as correlation within hospitals was not taken into account. Secondly, comparisons were made within (intervention and control) groups before and after the intervention but effects in the experimental and control groups were not directly compared. Re-analysis was, however, only possible for the outcome on preparedness for resuscitation - where baseline levels of performance ('mean practice scores') were comparable (intervention group 7.0 ± 4.08 versus control group 7.21 ± 4.51). The re-analysis involved comparison of intervention and control groups through computation of a mean difference using the reported standard deviations to estimate standard errors. To account for clustering we assumed an intra-cluster correlation coefficient (ICC) of 0.015 (with a design effect of 1.129), based on published data ([Rowe 2002](#)). The recalculated P value was annotated with 're-analysed'.

RESULTS

Description of studies

See: [Characteristics of included studies](#); [Characteristics of excluded studies](#).

See: [Characteristics of included studies](#); [Characteristics of excluded studies](#).

In total, we identified 2480 references from both the electronic and supplementary searches. No ongoing studies were identified. Not all the identified articles were published in English: 137 titles/abstracts (from the LILAC database) were in Portuguese and their titles were translated to English. We excluded 2334 irrelevant articles following a review of all the titles and abstracts. Reasons for exclusion included: inadequate study designs, inappropriate interventions/outcomes, enrolment of non-qualified healthcare providers, assessment of simulated practices outside practice settings, letters to the editor, commentaries, review articles, guidelines, non-paediatric studies, etc. We retrieved the full texts of 146 papers for further eligibility assessment. From these, we identified

eight studies as potentially meeting the review inclusion criteria. We excluded six of these studies for a variety of reasons following a detailed assessment (see [Characteristics of excluded studies](#)): [Bryce 2005](#), a non-randomised controlled study on health facility IMCI training, was excluded as the training intervention was combined and impossible to separate from concurrent district health strengthening activities such as skills reinforcement through supervised clinical practice (i.e. a complex intervention). [El-Arifeen 2004](#), a CRT on the effect of IMCI training on quality of care was excluded as data on referral rate (the appropriate health-worker response to an encounter with a seriously ill child, and our outcome of interest) were not reported for very ill children. We excluded another study ([Gouws 2004](#)) on the effect of IMCI on health worker antibiotic use as no baseline assessment of outcomes was performed in the IMCI trained and untrained groups. One intervention study ([Nadel 2000](#)) of periodic mock resuscitations combined with an 8-hour resuscitation course was excluded as it lacked a concurrent comparison group (i.e. used a historical control group). Two further studies were excluded as they enrolled only apparently well children ([Peltó 2004](#)) or those with mild ARI episodes ([Ochoa 1996](#)). Overall, we have considered two studies that met all the review inclusion criteria (see [Characteristics of included studies](#)). As a formal meta-analysis was not possible - given the small number of studies and differences in interventions (content, format, timing) and reported outcomes - a description of the included studies is provided below.

Both of the included studies were set in the delivery room/theatre in low-income countries (Kenya ([Opiyo 2008](#)), Sri Lanka ([Senarath 2007](#))). Both of the included studies were RCTs. The health professionals were nurses in one trial ([Opiyo 2008](#)) and mixed (doctors, nurses, midwives) in another ([Senarath 2007](#)). The targeted behaviours were process of initiating newborn resuscitation ([Opiyo 2008](#)) and general management/preparation and conduct of delivery care for newborns ([Senarath 2007](#)). The length of time during which the intervention was measured after initiation of intervention was 50 days in [Opiyo 2008](#), and three months in [Senarath 2007](#).

The number of experimental and control groups was balanced in [Opiyo 2008](#) but not in [Senarath 2007](#), where two hospitals were allocated to the intervention group and three hospitals to the control group. Both of the included studies were adequately powered (90%) for the primary outcomes. The unit of allocation in the included studies was healthcare professionals ($n = 83$) ([Opiyo 2008](#)), and hospitals ($n = 5$) ([Senarath 2007](#)). Both dichotomous (for example proportion of adequate resuscitation steps, proportion of newborns with undesirable health events) and continuous outcomes (for example frequency of harmful practices, mean scores of ENC practices) were considered in the included studies. None of the studies included information on the impact of the interventions on healthcare costs or resource utilization.

The first study ([Opiyo 2008](#)) was a RCT to determine if a simple one day newborn resuscitation training (NRT) alters initial

health worker resuscitation practices in a public hospital setting in Kenya. The intervention was a 1-day newborn resuscitation course adapted from the approach of the UK Resuscitation Council. The course teaches an A(Airway), B(Breathing), and C(Circulation) approach to resuscitation laying down a clear step by step strategy for the first minutes of resuscitation at birth. The teaching strategy was comprised of focused lectures and practical scenario sessions using infant manikins. Candidates were provided with a simple instruction manual two weeks before the training for self-learning. Health workers were randomly allocated to receive early training (n = 28) or late training (the control group, n = 55). Data were collected on 97 and 115 resuscitation episodes over 7 weeks after early training in the intervention and control groups respectively. The second study (Senarath 2007) was a RCT (with random allocation to groups) to evaluate the effectiveness of training for care providers on practice of essential newborn care in hospitals in Sri Lanka. The intervention was a 4-day training program on essential newborn care based on the WHO *Training Modules on Essential Newborn Care and Breastfeeding*. Additionally, participants were provided with teaching aids on *Newborn Care* (adapted from the National Neonatology Forum India) and *Resuscitation of the Newborn* (adapted from the Resuscitation Council (UK)). The teaching strategies involved lecture discussions, demonstrations, hands-on training, practical assignments, and small group discussions. Hospitals were randomly assigned to either the intervention group (n = 2 hospitals) or control group (n = 3 hospitals). The main sample for data collection by exit interview included 446 mother-newborn pairs pre-intervention and 446 post-intervention (223 each in intervention and control groups). These exit interview data were not relevant to the topic of this review. Direct observations of delivery practices were however made on a sub-sample consisting of 96 participants (48 before and 48 after the intervention). Post-intervention data collection commenced three months after the intervention.

Risk of bias in included studies

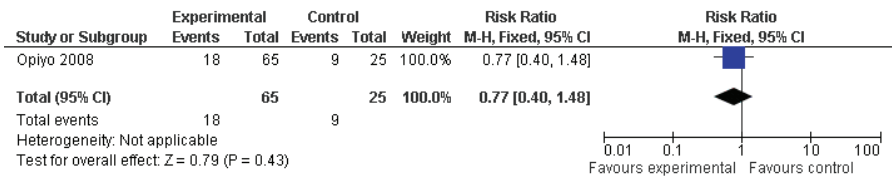
Both the included studies were of inadequate quality (high risk of bias) (see [Risk of bias in included studies](#)). In Opiyo 2008, blinding of outcome assessors and follow-up of providers was done, while allocation sequence generation, concealment, baseline measurement (of primary outcome), reporting of the reliability of outcome measures, and protection against contamination were not clear. In Senarath 2007, random allocation was adequately concealed, there was complete reporting of outcome data, and the study was adequately protected against contamination and selective outcome reporting. However, allocation sequence generation was unclear, and there were baseline differences in appropriate essential newborn care practices, and in the characteristics of study and control providers. Also, outcomes of interest were not assessed blindly and the presence of a 'unit of analysis error' could have contributed to additional risk of bias.

Effects of interventions

See: [Summary of findings for the main comparison](#) [Summary of Findings table](#)

In Opiyo 2008, trained providers demonstrated a higher proportion of adequate initial resuscitation steps compared to the control group (trained 66% versus control 27%; risk ratio 2.45, 95% confidence interval (CI) 1.75 to 3.42, P < 0.001, adjusted for clustering). In addition, there was a statistically significant reduction in the frequency of inappropriate and potentially harmful practices per resuscitation in the trained group (trained 0.53 versus control 0.92, mean difference 0.40, 95% CI 0.13 to 0.66, P = 0.004). Group comparison for the overall mortality in all resuscitation episodes (reported but not a stated primary outcome) showed no statistically significant differences between the groups (trained 0.28 (18/65), 95% CI 0.17 to 0.40; control 0.25 (9/25), 0.12 to 0.42, P = 0.77) (Figure 1).

Figure 1. Forest plot of comparison: 2 Opiyo 2008, outcome: 2.1 Mortality.



In [Senarath 2007](#), assessment of breathing of the newborn at birth and four out the five components of ENC practices improved in the intervention group three months after the intervention. Apart from the outcome on preparedness for resuscitation (see section on data synthesis above), it was not possible to re-analyse the data on other outcomes of interest. In the re-analysis accounted for clustering, ENC course was associated with a significant improvement in resuscitation preparedness (mean difference, MD 8.83, 95% CI 6.41 to 11.25, P value (re-analysed) < 0.001) ([Figure 2](#), [Figure 3](#)).

Figure 2. Forest plot of comparison: I Senarath 2007, outcome: I.1 Practice of preparedness of resuscitation.

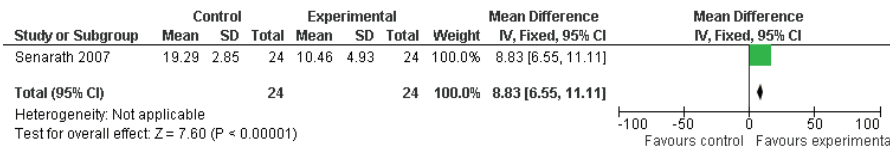
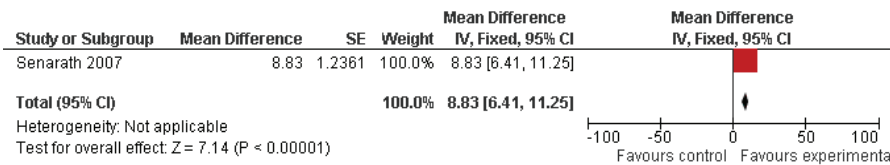


Figure 3. Forest plot of comparison: I Senarath 2007, outcome: I.2 Preparedness for resuscitation - adjusted for clustering.



DISCUSSION

This review found few well-conducted studies on the impact of neonatal or paediatric in-service training aimed at improving care for the seriously ill newborn or child. Limited evidence from the two included studies suggests a beneficial effect in the following outcomes: performance of initial resuscitation practices and reduction in the frequency of inappropriate practices ([Opiyo 2008](#)) in the short-term, and delivery room newborn care practices ([Senarath 2007](#)). We found no evidence of an effect on mortality, although the only study that reported this outcome was under-

powered to detect a mortality effect. Even though both the included studies reported positive performance outcomes following successful training, a generalisable evidence of effectiveness cannot be inferred - given the differences in interventions, outcomes, clinical settings and weaknesses in the study methods. These results are therefore intended for descriptive purposes only.

The common trend of benefit in the included studies should be interpreted with caution. First, in the study by [Opiyo 2008](#), assessments were conducted immediately following training for a short period of 50 days. Instantaneous improvement in performance would therefore have been expected. Clinical skills have however been shown to decay over time, with as much as a 50%

reduction in performance (as assessed in classroom simulations) within six months of intense training (McKenna 1985). Thus, evaluation of potential interaction effects of training over-time would have produced more reliable results. Conversely, the potential for a 'decay effect' underscores the need for periodic refresher trainings as a means to maintaining optimal performance especially given the infrequent nature of emergency care. Second, in Senarath 2007, baseline performance of newborn care practices was relatively high in both the intervention and control groups. Thus, the narrow 'performance improvement gap' could have limited possible demonstration of a real impact of the ENC program (i.e. possible 'ceiling effect'). Third, training coverage was low in Opiyo 2008 and unclear in Senarath 2007. Saturation training to the level of that reported in one excluded study (El-Arifeen 2004 (94%)) can potentially create a 'herd effect' on provider practices. Thus, possible mediation of reported effects by differences in levels of training coverage cannot be excluded.

The lengths of the considered training interventions varied: 1-day NRT in Opiyo 2008, and 4-day ENC course in Senarath 2007. Apart from the clear effect on costs, there is some evidence that the duration of training courses could influence their effectiveness: one related review (Rowe 2008) (n = 2 studies) which compared the standard IMCI training (duration \geq 11 days) to shortened training (5-11 days) reported marginal effectiveness of the standard in-service IMCI training course over the shortened training. In the same review, the effect of IMCI training over time was mixed with some analyses indicating increased effect with time, while others showed decreasing or no effect. In the current review, the length of follow-up period following training was relatively short (50 days) in Opiyo 2008 and three months in Senarath 2007. Thus, no reliable inference could be made regarding the magnitude of training effect over time. Still, to take account of the potential deterioration of clinical skills over time, it is recommended that evaluations of educational interventions include a sufficient length of follow-up period following the intervention. The effect of training could vary depending on the susceptibility of the targeted behaviour to the training intervention. Some behaviours (such as performance of inappropriate practices e.g. holding the baby upside down during resuscitation) are easier to change than others (such as correct performance of all resuscitation steps). In Opiyo 2008, the teaching strategy consisted of focused lectures and practical scenario sessions using an infant manikin, while in Senarath 2007, the strategy involved lecture discussions, demonstrations, hands-on training, practical assignments, and small group discussions. The content and format of in-service training courses could influence their effectiveness - in one Cochrane review on the effects of educational meetings on professional practice and healthcare outcomes (Forsetlund 2009), combined lectures and small group discussions appeared to be more effective. The outlined possible mediators of training effects add to the difficulty in deriving even a qualitative interpretation of the presented evidence.

The limited evidence available can be attributed to a number of factors: First, a significant number of studies were excluded on the basis of inadequate designs (e.g. lack of concurrent controls, use of historical controls, retrospective surveys, naturalistic designs, etc). Thus, the available evidence is mainly of poor quality with unreliable findings. Second, the lack of rigorous trials could also be attributed to design and ethical challenges inherent in the evaluation of educational interventions. Such desirable attributes as protection against contamination cannot practically be achieved within routine practice settings. Random assignment of healthcare providers and already vulnerable populations of infants to a control arm and observation of practices performed by untrained providers clearly raises ethical concerns. Third, effective sample sizes will always be hard to achieve for example severe illness episodes and resuscitation events remain relatively uncommon events in most clinical settings. Thus, large pragmatic multi-centre studies with prolonged observation periods would be needed to sufficiently demonstrate plausible changes in provider performance and ideally mortality. Apart from the clear logistical and cost implications, such trials would have to contend with the difficulty in securing the attendance and continued availability and participation of health workers. A possible optimal design to deal with the above tension between the need for high quality randomised evidence of effectiveness of emergency care courses and the highlighted ethical and practical constraints would be a pragmatic cluster-randomised trial with process evaluations to facilitate a better understanding of the determinants of actual practice (Elie 2007).

The findings of this review, in common with previous related reviews (Jabbour 1996, Rowe 2008), demonstrate the sparse evidence base for the impact of neonatal and paediatric courses on care of the seriously ill newborn or child. None of the included studies considered training programme development or implementation costs and thus any consideration of costs and benefits is impossible. While courses with a broader scope may have a broader range of benefits too there is limited evidence of effectiveness of emergency care courses. However, these courses continue to be popular within ministries of health and healthcare institutions, and are increasingly being promoted by influential groups such as the WHO and its partners. Before these become the standard of care, making them even more difficult to evaluate, evidence of their ability at least to change health worker practices and ideally to reduce mortality are required.

AUTHORS' CONCLUSIONS

Implications for practice

Studies included in this review do not provide a definitive evidence of effectiveness of in-service neonatal and paediatric courses in the emergency care setting. Additionally, despite the weak but positive evidence of benefit, it is still uncertain whether such in-service training, compared to alternative interventions, improves

outcomes at reasonable costs. The current findings cannot therefore be used to inform decisions on whether to invest in in-service emergency care training as opposed to other alternatives to improving the survival of seriously ill newborns or children.

Implications for research

Rigorous trials (with appropriate controls and adequate randomisation procedures) evaluating the impact of refresher emergency care training on long-term outcomes (professional practices and patient outcomes) are needed (given the current uncertainty on how long short-term benefits are retained, particularly in settings where they are used relatively infrequently). Such trials should: 1) involve direct head-to-head comparisons of courses with varied lengths (such as 1-day courses versus 4-day courses); 2) aim to include seriously ill newborns (in out-patient settings) and children (in both out-patient and hospital settings); and 3) include data on resources and cost of training implementation (to optimise appropriate policy decisions regarding which interventions to invest in). To facilitate replication, the studies should also provide sufficient detail regarding their content (e.g. need for equipment, teamwork

and format (e.g. small group interactive versus lectures, hands-on skills with dummies).

ACKNOWLEDGEMENTS

Newton Opiyo was awarded a Reviews for Africa Programme Fellowship (www.mrc.ac.za/cochrane/rap.htm), funded by a grant from The Nuffield Commonwealth Programme, through The Nuffield Foundation.

We would like to thank Marit Johansen and Kjetil Olsen for their assistance with the searches and the acquisition of full text articles. We would also like to thank Andy Oxman and Sasha Shepperd for their inputs at various stages of the review, Alexander Rowe for his helpful comments on the review manuscript, and Susan Babigumira for her editorial assistance.

Mike English is funded by a Wellcome Trust Senior Fellowship (#076827) and Newton Opiyo is supported by funding from a Wellcome Trust Strategic Award (#084538).

REFERENCES

References to studies included in this review

Opiyo 2008 *[published data only]*

Opiyo N, Were F, Govedi F, Fegan G, Wasunna A, English M. Effect of newborn resuscitation training on health worker practices in Pumwani Hospital, Kenya. *PLoS ONE* 2008;**13**(3(2)):e1599.

Senarath 2007 *[published data only]*

Senarath U, Fernando DN, Rodrigo I. Effect of training for care providers on practice of essential newborn care in hospitals in Sri Lanka. *Journal of Obstetric, Gynecologic and Neonatal Nursing* 2007;**36**(6):531–41.

References to studies excluded from this review

Bryce 2005 *[published data only]*

Bryce J, Gouws E, Adam T, Black RE, Schellenberg JA, Manzi F, et al. Improving quality and efficiency of facility-based child health care through Integrated Management of Childhood Illness in Tanzania. *Health Policy and Planning* 2005;**20**(Suppl 1):i69–i76.

El-Arifeen 2004 *[published data only]*

El Arifeen S, Blum LS, Hoque DM, Chowdhury EK, Khan R, Black RE, et al. Integrated Management of Childhood Illness (IMCI) in Bangladesh: early findings from a cluster-randomised study. *Lancet* 2004;**364**(9445):1595–602.

Gouws 2004 *[published data only]*

Gouws E, Bryce J, Habicht JP, Amaral J, Pariyo G, Schellenberg JA, et al. Improving antimicrobial use among health workers in first-level facilities: results from the multi-country evaluation of the Integrated Management of Childhood Illness strategy. *Bulletin of the World Health Organization* 2004;**82**(7):509–15.

Nadel 2000 *[published data only]*

Nadel FM, Lavelle JM, Fein JA, Giardino AP, Decker JM, Durbin DR. Teaching resuscitation to pediatric residents: the effects of an intervention. *Archives of Pediatrics and Adolescent Medicine* 2000;**154**(10):1049–54.

Ochoa 1996 *[published data only]*

González Ochoa E, Armas Pérez L, Bravo González JR, Cabrales Escobar J, Rosales Corrales R, Abreu Suárez G. Prescription of antibiotics for mild acute respiratory infections in children. *Bulletin of the Pan American Health Organisation* 1996;**30**(2):106–17.

Pelto 2004 *[published data only]*

Pelto GH, Santos I, Gonçalves H, Victora C, Martinez J, Habicht JP. Nutrition counseling training changes physician behavior and improves caregiver knowledge acquisition. *The Journal of Nutrition* 2004;**134**(2):357–62.

Additional references

Baskett 2005

Baskett PJ, Nolan JP, Handley A, Soar J, Biarent D, Richmond S. European Resuscitation Council. European resuscitation council guidelines for resuscitation 2005. Section 9. Principles of training in resuscitation. *Resuscitation* 2005;**67**S1:S181–9.

Berkley 2005

Berkley JA, Maitland K, Mwangi I, Ngetsa C, Mwarumba S, Lowe BS, et al. Use of clinical syndromes to target antibiotic prescribing in seriously ill children in malaria endemic area: observational study. *BMJ* 2005;**330**(7498):995.

Elie 2007

Akl EA, Treweek S, Foy R, Francis J, Oxman AD, for the ReBEQI group. NorthStar, a support tool for the design and evaluation of quality improvement interventions in healthcare. *Implementation Science* 2007;**2**:19 doi:10.1186/1748-5908-2-19.

EPOC 2007

EPOC 2007. EPOC. Resources for review authors- reviewer tools. www.epoc.cochrane.org/en/index.html (accessed 19 December 2007).

Forsberg 2007

Forsberg BC, Petzold MG, Thompson G, Allebeck P. Diarrhoea case management in low and middle-income countries-an unfinished agenda. *Bulletin of the World Health Organization* 2007;**85**(1):42-8.

Forsetlund 2009

Forsetlund L, Bjørndal A, Rashidian A, Jamtvedt G, O'Brien MA, Wolf F, et al. Continuing education meetings and workshops: effects on professional practice and health care outcomes. *Cochrane Database of Systematic Reviews* 2009, Issue 2. [DOI: 10.1002/14651858.CD003030.pub2]

Gove 1999

Gove S, Tamburlini G, Molyneux E, Whitesell P, Campbell H. Development and technical basis of simplified guidelines for emergency triage assessment and treatment in developing countries. WHO Integrated Management of Childhood Illness (IMCI) Referral Care Project. *Archives of Disease in Childhood* 1999;**81**(6):473-7.

Higgins 2006

Higgins JPT, Green S, (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 4.2.6 [Updated September 2006]. The Cochrane Collaboration, 2006. Available from www.cochrane-handbook.org.

Irimu 2008

Irimu G, Wamae A, Wasunna A, Were F, Ntoburi S, Opiyo N, et al. Developing and introducing evidence based clinical practice guidelines for serious illness in Kenya. *Archives of Disease in Childhood* 2008;**93**(9):799-804.

Jabbour 1996

Jabbour M, Osmond MH, Klassen TP. Life support courses: are they effective?. *Annals of Emergency Medicine* 1996;**28**:690-8.

Jewkes 2003

Jewkes F, Phillips B. Resuscitation training of paediatricians. *Archives of Disease in Childhood* 2003;**88**:118-21.

McKenna 1985

McKenna S, Glendon A. Occupational first aid training: decay in cardiopulmonary resuscitation (CPR) skills. *Journal of occupational psychology* 1985;**58**:109-17.

Mello 2003

Mello MJ, dos Santos ER, Buzzati R, et al. Pregnancy, childbirth, postpartum and newborn care guide for essential practice. In: Integrated management of childbirth. *Geneva: World Health Organization* 2003:1-178.

Molyneux 2006

Molyneux E, Ahmad S, Robertson A. Improved triage and emergency care for children reduces inpatient mortality in a resource-constrained setting. *Bulletin of the World Health Organization* 2006;**84**(4):314-9.

Nolan 2001

Nolan T, Angus P, Cunha AJ, Muhe L, Qazi S, Simoes EA, et al. Quality of hospital care for seriously ill children in less-developed countries. *Lancet* 2001;**357**(9250):106-10.

Pio 2003

Pio 2003. Standard case management of pneumonia in children in developing countries: the cornerstone of the acute respiratory infection programme. *Bulletin of the World Health Organization* 2003;**81**(4):298-300.

Raupp 2007

Raupp P, McCutcheon C. Neonatal resuscitation-An analysis of the transatlantic divide. *Resuscitation* 2007;**75**(2):345-9.

Rowe 2002

Rowe 2002. Design effects and intraclass correlation coefficients from a health facility cluster survey in Benin. *International Journal for Quality in Health Care* 2002;**14**(6):521-523.

Rowe 2008

Rowe AK, Rowe SY, Holloway KA, Ivanovska I, Muhe L, Lambrechts T. A systematic review of the effectiveness of shortening Integrated Management of Childhood Illness guidelines training: final report. World Health Organization 2008.

Sazawal 2001

Sazawal S, Black RE, for the Pneumonia Case Management Trials Group. Effect of pneumonia case management on mortality in neonates, infants, and preschool children: a meta-analysis of community-based trials. *Lancet Infectious Diseases* 2001;**3**(9):547-56.

Sethi 2001

Sethi D, Kwan I, Kelly AM, Roberts I, Bunn F. Advanced trauma life support training for ambulance crews. *Cochrane Database of Systematic Reviews* 2001, Issue 2. [DOI: 10.1002/14651858.CD003109]

Shakiba 2003

Shakiba H, Dinesh S, Anne MK. Advanced trauma life support training for hospital staff. *Cochrane Database of Systematic Reviews* 2003, Issue 3. [DOI: %3Chtml%3E%3Cbody id=%22body%22%3E10.1002/14651858.CD004173.pub2%3C/body%3E%3C/html%3E]

Tulloch 1999

Tulloch J. Integrated approach to child health in developing countries. *Lancet* 1999;**354** (Suppl. 2):S1116-20.

WHO 2002

WHO 2002. World Health Organization. Training course on the management of severe malnutrition. *Geneva: WHO* 2002.

Wello 2005

World Health Organization. The World Health Report: 2005: make every mother and child count. Geneva: World Health Organization 2005.

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies [ordered by study ID]

Opiyo 2008

Methods	Randomised controlled trial (RCT)
Participants	Nurses Country: Kenya Phase 1: 83 nurses (28 intervention, 55 control) 97 practices in intervention group; 115 practices in control group Type of targeted behaviour - process of initiating newborn resuscitation
Interventions	Newborn resuscitation training (NRT) Duration of intervention - 1 day Co-intervention - self learning instruction manual provided to participants 2 weeks prior to training Control: standard practice Length of time during which intervention was measured after initiation of intervention - 50 days (phase 1)
Outcomes	Proportion of appropriate initial resuscitation steps Frequency of inappropriate/harmful practices Mortality
Notes	No difference between comparison groups in age and number of years worked Review authors also authors in the study (see conflict of interest statement) Primary analysis based on phase 1 data only Overall study quality: high risk of bias

Risk of bias

Item	Authors' judgement	Description
Adequate sequence generation?	Unclear	'...our intention was to randomise staff, stratified by place of work...'. The specific random approach was however not specified.
Allocation concealment?	Unclear	Health worker used as the unit of clustering but the random process incompletely reported.
Blinding? All outcomes	Yes	'The observers were blind to the training status of the health workers and were instructed not to try to ascertain health workers' training status...'
Contamination? All outcomes	Unclear	'We cannot exclude the possibility of cross-group contamination...'
Baseline characteristics?	Yes	'There were no significant differences in the ages...and years of experience between the groups...'

Opiyo 2008 (Continued)

Incomplete data?	Yes	'32 allocated to intervention....28 providers observed', '58 allocated to control...55 providers observed'
Other risks of bias?	Yes	

Senarath 2007

Methods	Randomised controlled trial
Participants	Doctors, nurses, midwives Country: Sri Lanka 110 participants (59 intervention, 61 control) Type of targeted behaviour: general management - preparation and conduct of delivery care for newborn
Interventions	Essential newborn care (ENC) course Duration of intervention: 4 days Co-interventions: none Control: standard practice
Outcomes	Practices of essential newborn care at delivery
Notes	Reported results restricted to observation data only Length of time during which intervention was measured after initiation of intervention - 3 months 'Unit of analysis error present': unit of randomisation - hospitals; unit of analysis - observed delivery practices. Also, effects in experimental and control groups not directly compared. Overall study quality: high risk of bias

Risk of bias

Item	Authors' judgement	Description
Allocation concealment?	Yes	'...hospitals were randomly assigned to either of two groups...the intervention group...and control groups...'
Baseline outcome measurement? All outcomes	No	'There were some differences in the baseline level of practices between intervention and control groups...'
Blinding? All outcomes	No	'The principal investigator made observations in labor room...'
Contamination? All outcomes	Yes	Hospitals were randomly allocated to comparison groups, and it is unlikely that the control group received ENC training

Incomplete data?	Yes	'...participants (48 before and after the intervention) was selected...'. Effect of intervention on observed practices reported for 48 participants before and after the intervention'
Other risks of bias?	No	'Unit of analysis error'

Characteristics of excluded studies *[ordered by study ID]*

Bryce 2005	Non-randomised design with concurrent health system strengthening activities (complex intervention)
El-Arifeen 2004	Data on referral rate for very ill children (outcome of interest) not reported
Gouws 2004	No baseline assessment of outcomes in IMCI trained and untrained groups
Nadel 2000	Study has a historical group only and used mock scenarios to assess practice
Ochoa 1996	Study did not include seriously ill children (only considered mild ARI episodes)
Pelto 2004	Study focused on an IMCI derived nutrition counselling protocol in apparently well children

DATA AND ANALYSES

Comparison 1. Opiyo 2008

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Mortality	1	90	Risk Ratio (M-H, Fixed, 95% CI)	0.77 [0.40, 1.48]

Comparison 2. Senarath 2007

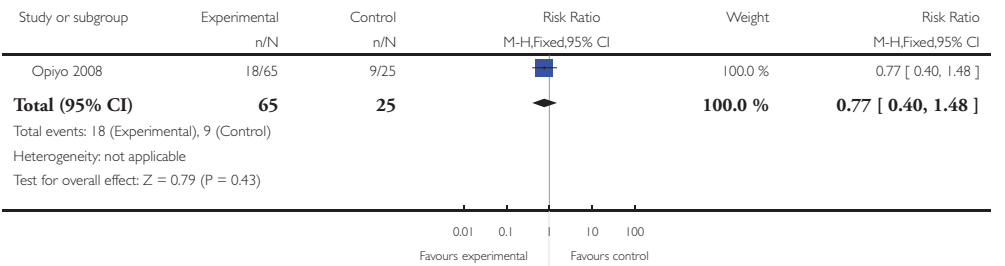
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Practice of preparedness of resuscitation	1	48	Mean Difference (IV, Fixed, 95% CI)	8.83 [6.55, 11.11]
2 Preparedness for resuscitation - adjusted for clustering	1		Mean Difference (Fixed, 95% CI)	8.83 [6.41, 11.25]

Analysis 1.1. Comparison 1 Opiyo 2008, Outcome 1 Mortality.

Review: In-service training for health professionals to improve care of the seriously ill newborn or child in low and middle-income countries (Review)

Comparison: 1 Opiyo 2008

Outcome: 1 Mortality

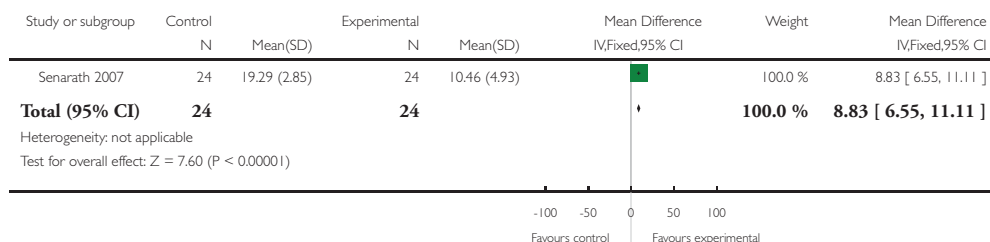


Analysis 2.1. Comparison 2 Senarath 2007, Outcome 1 Practice of preparedness of resuscitation.

Review: In-service training for health professionals to improve care of the seriously ill newborn or child in low and middle-income countries (Review)

Comparison: 2 Senarath 2007

Outcome: 1 Practice of preparedness of resuscitation

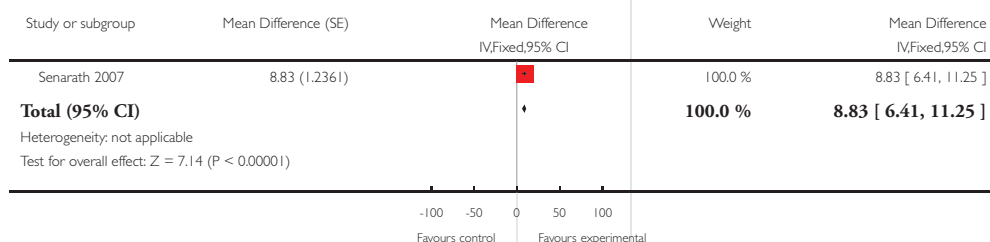


Analysis 2.2. Comparison 2 Senarath 2007, Outcome 2 Preparedness for resuscitation - adjusted for clustering.

Review: In-service training for health professionals to improve care of the seriously ill newborn or child in low and middle-income countries (Review)

Comparison: 2 Senarath 2007

Outcome: 2 Preparedness for resuscitation - adjusted for clustering



APPENDICES

Appendix I. Detailed search strategy: CENTRAL / EPOC Register (up to May Week I 2009)

Search terms

- #1 MeSH descriptor Inservice Training explode all trees
- #2 MeSH descriptor Health Personnel explode all trees with qualifier: ED
- #3 MeSH descriptor Internship and Residency, this term only
- #4 (staff or employee* or clinician* or physician? or nurs* or midwife* or midwiv* or pharmacist* or specialist* or practitioner* or dietician* or nutritionist*) NEXT (train* or course* or development or education or teach*):ti or (staff or employee* or clinician* or physician? or nurs* or midwife* or midwiv* or pharmacist* or specialist* or practitioner* or dietician* or nutritionist*) NEXT (train* or course* or development or education or teach*):ab
- #5 (inservice or in NEXT service or life NEXT support) NEAR/2 (train* or course* or development or education or teach*):ti or (inservice or in NEXT service or life NEXT support) NEAR/2 (train* or course* or development or education or teach*):ab
- #6 “on the job training”:ti or “on the job training”:ab
- #7 (#1 OR #2 OR #3 OR #4 OR #5 OR #6)
- #8 MeSH descriptor Case Management, this term only
- #9 MeSH descriptor Critical Care explode all trees
- #10 MeSH descriptor Life Support Care, this term only
- #11 MeSH descriptor Critical Illness, this term only
- #12 MeSH descriptor Acute Disease, this term only
- #13 MeSH descriptor Emergency Medical Services explode all trees
- #14 MeSH descriptor Emergency Medicine, this term only
- #15 MeSH descriptor Emergency Treatment explode all trees
- #16 MeSH descriptor Emergency Nursing, this term only
- #17 (case NEXT management):ti or (case NEXT management):ab
- #18 (emergency NEXT triage*):ti or (emergency NEXT triage*):ab
- #19 (life NEXT support):ti or (life NEXT support):ab
- #20 (resuscitation):ti or (resuscitation):ab
- #21 (first NEXT aid):ti or (first NEXT aid):ab
- #22 (referral or urgent) NEAR/2 care:ti or (referral or urgent) NEAR/2 care:ab
- #23 (critical* or emergency or intensive or serious* or sever* or acute*) NEAR/2 (care or ill or illness* or treatment or therap*):ti or (critical* or emergency or intensive or serious* or sever* or acute*) NEAR/2 (care or ill or illness* or treatment or therap*):ab
- #24 (#8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23)
- #25 MeSH descriptor Child explode all trees
- #26 MeSH descriptor Infant explode all trees
- #27 MeSH descriptor Child Care explode all trees
- #28 MeSH descriptor Pediatrics explode all trees
- #29 MeSH descriptor Pediatric Nursing explode all trees
- #30 MeSH descriptor Perinatal Care, this term only
- #31 (child* or infant* or pediatric or paediatric or newborn* or neonat* or baby or babies or kid* or toddler*):ti or (child* or infant* or pediatric or paediatric or newborn* or neonat* or baby or babies or kid* or toddler*):ab
- #32 (#25 OR #26 OR #27 OR #28 OR #29 OR #30 OR #31)
- #33 MeSH descriptor Child Care explode all trees with qualifier: ED
- #34 MeSH descriptor Pediatrics explode all trees with qualifier: ED
- #35 MeSH descriptor Pediatric Nursing explode all trees with qualifier: ED
- #36 (#33 OR #34 OR #35)
- #37 MeSH descriptor Critical Care explode all trees with qualifier: ED
- #38 MeSH descriptor Life Support Care, this term only with qualifier: ED
- #39 MeSH descriptor Emergency Medical Services explode all trees with qualifier: ED
- #40 MeSH descriptor Emergency Medicine, this term only with qualifier: ED
- #41 MeSH descriptor Emergency Treatment explode all trees with qualifier: ED

#42 MeSH descriptor Emergency Nursing, this term only with qualifier: ED
 #43 (#37 OR #38 OR #39 OR #40 OR #41 OR #42)
 #44 MeSH descriptor Intensive Care, Neonatal, this term only
 #45 MeSH descriptor Diarrhea, Infantile, this term only
 #46 MeSH descriptor Infant, Newborn, Diseases explode all trees
 #47 "Acute Respiratory Infection" or "Acute Respiratory Infections":ti or "Acute Respiratory Infection" or "Acute Respiratory Infections":ab
 #48 (#44 OR #45 OR #46 OR #47)
 #49 "Control of Diarrheal Disease" or "Control of Diarrheal Diseases":ti or "Control of Diarrheal Disease" or "Control of Diarrheal Diseases":ab
 #50 "Neonatal Resuscitation Program" or "Neonatal Resuscitation Programs":ti or "Neonatal Resuscitation Program" or "Neonatal Resuscitation Programs":ab
 #51 "Essential Newborn Care":ti or "Essential Newborn Care":ab
 #52 "Integrated Management of Childhood Illness":ti or "Integrated Management of Childhood Illness":ab
 #53 (#49 OR #50 OR #51 OR #52)
 #54 (#7 AND #24 AND #32)
 #55 (#24 AND #36)
 #56 (#32 AND #43)

Appendix 2. Detailed search strategy: Ovid MEDLINE(R) 1950 to May Week 1 2009

Search terms

1	exp Inservice Training/
2	exp Health Personnel/ed [Education]
3	"Internship and Residency"/
4	((staff or employee? or clinician? or physician? or nurs\$ or midwif\$ or midwiv\$ or pharmacist? or specialist? or practitioner? or dietician? or nutritionist?) adj (train\$ or course? or development or education or teach\$)).tw.
5	((inservice or in-service or life support) adj2 (train\$ or course? or development or education or teach\$)).tw.
6	on the job training.tw.
7	or/1-6
8	Case Management/
9	exp Critical Care/
10	Life Support Care/
11	Critical Illness/
12	Acute Disease/
13	exp Emergency Medical Services/

(Continued)

14	Emergency Medicine/
15	exp Emergency Treatment/
16	Emergency Nursing/
17	case management.tw.
18	emergency triage?.tw.
19	life support.tw.
20	resuscitation.tw.
21	first aid.tw.
22	((referral or urgent) adj2 care).tw.
23	((critical\$ or emergency or intensive or serious\$ or sever\$ or acute\$) adj2 (care or ill or illness\$ or treatment or therap\$)).tw.
24	or/8-23
25	exp Child/
26	exp Infant/
27	exp Child Care/
28	Pediatrics/
29	Neonatology/
30	Perinatology/
31	Pediatric Nursing/
32	Perinatal Care/
33	Neonatal Nursing/
34	(child\$ or infant? or pediatric or paediatric or newborn? or neonat\$ or baby or babies or kid? or toddler?).tw.
35	or/25-34
36	exp Child Care/ed [Education]
37	Pediatrics/ed [Education]

(Continued)

38	Neonatology/ed [Education]
39	Perinatology/ed [Education]
40	Pediatric Nursing/ed [Education]
41	Neonatal Nursing/ed [Education]
42	or/36-41
43	exp Critical Care/ed [Education]
44	Life Support Care/ed [Education]
45	exp Emergency Medical Services/ed [Education]
46	Emergency Medicine/ed [Education]
47	exp Emergency Treatment/ed [Education]
48	Emergency Nursing/ed [Education]
49	or/43-48
50	Intensive Care, Neonatal/
51	Diarrhea, Infantile/
52	Acute Respiratory Infection?.tw.
53	or/50-52
54	exp Infant, Newborn, Diseases/
55	Control of Diarrheal Disease?.tw.
56	Neonatal Resuscitation Program?.tw.
57	Essential Newborn Care.tw.
58	Integrated Management of Childhood Illness.tw.
59	or/55-58
60	7 and 24 and 35
61	24 and 42

(Continued)

62	35 and 49
63	7 and 53
64	7 and 24 and 54
65	or/59-64
66	Randomized Controlled Trial.pt.
67	random\$.tw.
68	intervention?.tw.
69	control\$.tw.
70	evaluat\$.tw.
71	effect\$.tw.
72	or/66-71
73	Animal/
74	Human/
75	73 not (73 and 74)
76	letter.pt.
77	editorial.pt.
78	comment.pt.
79	72 not (75 or 76 or 77 or 78)
80	65 and 79

Appendix 3. Detailed search strategy: EMBASE (1980 to May Week 1 2009)

Search terms

1	In Service Training/
2	Staff Training/

(Continued)

3	Nurse Training/
4	Continuing Education/
5	Professional Development/
6	Medical Education/
7	Residency Education/
8	((staff or employee? or clinician? or physician? or nurs\$ or midwif\$ or midwiv\$ or pharmacist? or specialist? or practitioner? or dietician? or nutritionist?) adj (train\$ or course? or development or education or teach\$)).tw.
9	((inservice or in-service or life support) adj2 (train\$ or course? or development or education or teach\$)).tw.
10	on the job training.tw.
11	or/1-10
12	Case Management/
13	exp Intensive Care/
14	Critical Illness/
15	Disease Severity/
16	Acute Disease/
17	Injury Severity/
18	Emergency Medicine/
19	exp Emergency Treatment/
20	Emergency Nursing/
21	case management.tw.
22	emergency triage?.tw.
23	life support.tw.
24	resuscitation.tw.
25	first aid.tw.

(Continued)

26	((referral or urgent) adj2 care).tw.
27	((critical\$ or emergency or intensive or serious\$ or sever\$ or acute\$) adj2 (care or ill or illness\$ or treatment or therap\$)).tw.
28	or/12-27
29	exp Child/
30	exp Newborn/
31	exp Child Health Care/
32	exp Pediatrics/
33	exp Pediatric Nursing/
34	exp Postnatal Care/
35	Perinatal Care/
36	(child\$ or infant? or pediatric or paediatric or newborn? or neonat\$ or baby or babies or kid? or toddler?).tw.
37	or/29-36
38	Newborn Intensive Care/
39	Newborn Intensive Care Nursing/
40	Pediatric Intensive Care Nursing/
41	Pediatric Advanced Life Support/
42	Infantile Diarrhea/
43	Acute Respiratory Infection?.tw.
44	or/38-43
45	Emergency Medical Services Education/
46	exp Newborn Disease/
47	Control of Diarrheal Disease?.tw.
48	Neonatal Resuscitation Program?.tw.
49	Essential Newborn Care.tw.

(Continued)

50	Integrated Management of Childhood Illness.tw.
51	or/47-50
52	11 and 28 and 37
53	11 and 44
54	37 and 45
55	11 and 28 and 46
56	or/51-55
57	Randomized Controlled Trial/
58	Time Series Analysis/
59	random\$.tw.
60	experiment\$.tw.
61	(time adj series).tw.
62	(pre test or pretest or post test or posttest).tw.
63	impact.tw.
64	intervention?.tw.
65	chang\$.tw.
66	evaluat\$.tw.
67	effect?.tw.
68	compar\$.tw.
69	control\$.tw.
70	or/57-69
71	Nonhuman/
72	letter.pt.
73	editorial.pt.

(Continued)

74	70 not (71 or 72 or 73)
75	56 and 74

Appendix 4. Detailed search strategy: CINAHL (1982 to March Week 1 2008)

Search terms

1	exp Staff Development/
2	exp Health Personnel/ed [Education]
3	"Internship and Residency"/
4	((staff or employee? or clinician? or physician? or nurs\$ or midwif\$ or midwiv\$ or pharmacist? or specialist? or practitioner? or dietician? or nutritionist?) adj (train\$ or course? or development or education or teach\$)).tw.
5	((inservice or in-service or life support) adj2 (train\$ or course? or development or education or teach\$)).tw.
6	on the job training.tw.
7	or/1-6
8	Case Management/
9	exp Critical Care/
10	Life Support Care/
11	Critical Illness/
12	Acute Disease/
13	exp Emergency Medical Services/
14	Emergency Medicine/
15	First Aid/
16	exp Resuscitation/
17	exp Emergency Nursing/

(Continued)

18	case management.tw.
19	emergency triage?.tw.
20	life support.tw.
21	resuscitation.tw.
22	first aid.tw.
23	((referral or urgent) adj2 care).tw.
24	((critical\$ or emergency or intensive or serious\$ or sever\$ or acute\$) adj2 (care or ill or illness\$ or treatment or therap\$)).tw.
25	or/8-24
26	exp Child/
27	exp Child Care/
28	Child Health/
29	(child\$ or infant? or pediatric or paediatric or newborn? or neonat\$ or baby or babies or kid? or toddler?).tw.
30	Pediatrics/
31	Neonatology/
32	Perinatal Care/
33	Prenatal Care/
34	exp Pediatric Care/
35	exp Pediatric Nursing/
36	or/26-35
37	exp Child Care/ed [Education]
38	Child Health/ed [Education]
39	exp Pediatrics/ed [Education]
40	Perinatal Care/ed [Education]
41	Prenatal Care/ed [Education]

(Continued)

42	exp Pediatric Care/ed [Education]
43	exp Pediatric Nursing/ed [Education]
44	or/37-43
45	exp Critical Care/ed [Education]
46	Life Support Care/ed [Education]
47	exp Emergency Medical Services/ed [Education]
48	Emergency Medicine/ed [Education]
49	First Aid/ed [Education]
50	exp Resuscitation/ed [Education]
51	exp Emergency Nursing/ed [Education]
52	or/45-51
53	exp Intensive Care, Neonatal/
54	Acute Respiratory Infection?.tw.
55	or/53-54
56	exp Infant, Newborn, Diseases/
57	Control of Diarrheal Disease?.tw.
58	Neonatal Resuscitation Program?.tw.
59	Integrated Management of Childhood Illness.tw.
60	Pediatric Advanced Life Support/ed [Education]
61	or/57-60
62	7 and 25 and 36
63	25 and 44
64	36 and 52
65	7 and 55

(Continued)

66	7 and 25 and 56
67	or/61-66
68	Clinical Trial/
69	exp Pretest-Posttest Design/
70	exp Quasi-Experimental Studies/
71	Comparative Studies/
72	control\$.tw.
73	random\$.tw.
74	experiment\$.tw.
75	(time adj series).tw.
76	impact.tw.
77	intervention?.tw.
78	evaluat\$.tw.
79	effect\$.tw.
80	(pre test or pretest or post test or posttest).tw.
81	or/68-80
82	67 and 81

Appendix 5. Detailed search strategy: LILACS (up to October 2008)

Search terms

(inservice and training) or (inservice and course\$) or (inservice and workshop\$) or (inservice and education) or (inservice and program\$) or (capacitación and servicio) or (capacitação and serviço) [Palavras]

And

child or children or niño or criança or infant or infants or lactante or lactente or pediatric\$ or paediatric\$ or pediatria or pediatria or newborn or (recién and nacidos) or (recém and nascidos) or neonat\$ or baby or babies or kid or kids or toddler\$ [Palavras]

Appendix 6. Detailed search strategy: ERIC (up to October 2008)

Search terms

(DE=Inservice Education or DE=On the Job Training or TI=inservice training or TI=on the job training or TI=inservice course* or TI=inservice workshop* or TI=inservice education or TI=inservice program* or TI=in service training or TI=in service course* or TI=in service workshop* or TI=in service education or TI=in service program* or AB=inservice training or AB=on the job training or AB=inservice course* or AB=inservice workshop* or AB=inservice education or AB=inservice program* or AB=in service training or AB=in service course* or AB=in service workshop* or AB=in service education or AB=in service program*)

AND

(DE=Crisis Management or DE=Crisis Intervention or DE=Emergency Programs or DE=First Aid or TI=crisis management or TI=crisis intervention* or TI=emergency program* or AB=crisis management or AB= crisis intervention* or AB=emergency program* or TI=critical care or TI=critical* ill* or TI=critical treatment* or TI=critical therap* or TI=emergency care or TI=emergency ill* or TI=emergency treatment* or TI=emergency therap* or TI=intensive care or TI=intensive ill* or TI=intensive treatment* or TI=intensive therap* or TI=serious care or TI=serious ill* or TI=serious treatment* or TI=serious therap* or TI=sever* care or TI=sever* ill* or TI=sever* treatment* or TI=sever* therap* or TI=acute* care or TI=acute* ill* or TI=acute* treatment* or TI=acute* therap* or TI=first aid or TI=life support or TI=urgent care or TI=resuscitation or AB=critical care or AB=critical* ill* or AB=critical treatment* or AB=critical therap* or AB=emergency care or AB=emergency ill* or AB=emergency treatment* or AB=emergency therap* or AB=intensive care or AB=intensive ill* or AB=intensive treatment* or AB=intensive therap* or AB=serious care or AB=serious ill* or AB=serious treatment* or AB=serious therap* or AB=sever* care or AB=sever* ill* or AB=sever* treatment* or AB=sever* therap* or AB=acute* care or AB=acute* ill* or AB=acute* treatment* or AB=acute* therap* or AB=first aid or AB=life support or AB=urgent care or AB=resuscitation)

AND

(DE=Hospitalized Children or DE=children or DE=young children or DE=toddlers or DE=infants or DE=infant care or DE=premature infants or DE=neonates or DE=pediatrics or TI=child or TI=children or TI=infant or TI=infants or TI=pediatric* or TI=paediatric* or TI=newborn or TI=neonat* or TI=baby or TI=babies or TI=kid or TI=kids or TI=toddler* or AB=child or AB=children or AB=infant or AB=infants or AB=pediatric* or AB=paediatric* or AB=newborn or AB=neonat* or AB=baby or AB=babies or AB=kid or AB=kids or AB=toddler*)

Appendix 7. Detailed search strategy: WHOLIS (up to October 2008)

Search terms

words or phrase

inservice or 'in service' or 'on the job'

AND words or phrase

training or course\$ or education or workshop\$ or program\$

AND words or phrase

child\$ or infant\$ or pediatric\$ or paediatric\$ or newborn\$ or neonat\$ or baby or babies or kid or kids or toddler\$

WHAT'S NEW

Last assessed as up-to-date: 21 May 2009.

22 March 2010	Amended	Minor edits
---------------	---------	-------------

HISTORY

Protocol first published: Issue 2, 2008

Review first published: Issue 4, 2010

CONTRIBUTIONS OF AUTHORS

NO and ME wrote the protocol, screened records for eligibility, extracted data, assessed methodological quality of included studies, interpreted findings and wrote the review. NO prepared the first draft of the review. ME commented on the manuscript.

DECLARATIONS OF INTEREST

NO and ME are authors of one of the studies [Opiyo 2008](#) included in this review. The methodological quality of this study was also considered by another EPOC reviewer.

SOURCES OF SUPPORT

Internal sources

- KEMRI/Wellcome Trust Research Programme, Kenya.

External sources

- South African Cochrane Centre, South Africa.
- Cochrane Effective Practice and Organisation of Care Group (EPOC), Norway.

INDEX TERMS

Medical Subject Headings (MeSH)

*Developing Countries; Infant, Newborn; Inservice Training [*methods]; Neonatology [*education]; Perinatology [*education]; Quality of Health Care [*standards]; Randomized Controlled Trials as Topic

MeSH check words

Child; Humans

Comparison of alternative evidence summary and presentation formats in clinical guideline development: a mixed-method evaluation

Newton Opiyo^{a, b*}

Sasha Shepperd^c

Nyokabi Musila^a

Elizabeth Allen^d

Rachel Nyamai^e

Atle Fretheim^{b, f}

Mike English^{a, g}

^aKEMRI-Wellcome Trust Research Programme, P.O. Box 43640, Nairobi 00100, Kenya

^bInstitute of Health and Society, University of Oslo, P.O. Box 1130, Blindern, 0318, Oslo, Norway

^cDepartment of Public Health, University of Oxford, Headington, OX3 7LF, Oxford, UK

^dLondon School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

^eDivision of Paediatric, Ministry of Medical Services, P.O. Box 30016, Nairobi 00100, Nairobi, Kenya

^fNorwegian Knowledge Centre for Health Services, P.O. Box 7004 St. Olavs plass, N-0130, Oslo, Norway

^gDepartment of Paediatrics, University of Oxford, John Radcliffe Hospital, Headington, Oxford, OX3 9DU, UK

*** Corresponding author:**

Newton Opiyo

KEMRI-Wellcome Trust Research Programme, P.O. Box 43640, Nairobi 00100, Kenya

nopiyo@nairobi.kemri-wellcome.org

Word Count:	Abstract	247
	Main text	3,071

Keywords: clinical practice guidelines; knowledge translation; evidence synthesis; systematic reviews

Running title: Evidence summary formats in clinical guideline development

Abstract

Objective

To assess the usefulness of different formats for summarising and presenting evidence for use in clinical guideline development.

Methods

Healthcare professionals attending a one-week Kenyan, national guideline development workshop were randomly allocated to receive evidence packaged in three different formats: systematic reviews (SR) alone, systematic reviews with summary-of-findings tables (SR with SoF tables), and 'graded-entry' formats (a 'front-end' summary and a contextually framed narrative report plus the SR). The influence of format on ability to retrieve key outcome information, the primary outcome, was assessed using a written test. Interviews conducted within two months following completion of trial data collection explored panel members' views on the evidence summary formats and experiences with appraisal and use of research information.

Findings

Sixty-five trial participants completed the questionnaire. There were no differences between the comparison groups in the odds of correct responses to key clinical questions (adjusted odds ratios, SR with SoF tables versus SR alone: 0.59, 95% confidence interval, CI 0.32 to 1.07; 'graded-entry' format versus SR alone: 0.66, 95% CI 0.36 to 1.21). 'Graded-entry' formats were associated with a higher mean composite score for clarity and accessibility of information about the quality of evidence than SRs alone (adjusted mean difference 0.52, 95% CI: 0.06 to 0.99). Findings from interviews with 16 panelists indicated that short narrative evidence reports were preferred for the improved clarity of information presentation and ease of use.

Conclusion

'Graded-entry' evidence summary formats may improve clarity and accessibility of research evidence in clinical guideline development.

Study registration: The study is registered at <http://www.controlled-trials.com>, number ISRCTN05154264.

Evidence summary formats in clinical guideline development

What is already known on this topic

- Best formats for summarising and presenting evidence for use in clinical guideline development remain unclear.

What this study adds

- 'Graded-entry' evidence summary formats (a 'front-end' summary of key information and a contextually framed narrative report plus the full systematic review) may improve clarity and accessibility of research evidence in clinical guideline development.
- Providing a 'front-end' summary of key information linked to locally relevant factors that support implementation, and the full systematic review, may help those developing guidelines access and contextualise research evidence.

INTRODUCTION

Policy decisions about which interventions to implement, modify or withdraw from health care, should be informed by the best available evidence. While systematic reviews are a key source of evidence,¹⁻² technical jargon, and the time it takes to read and identify the key findings in a review, may deter healthcare decision makers from applying this evidence.³ Thus, alternative ways of summarising and presenting evidence have been developed to improve accessibility and use (Box 1). Such summaries need to be up to date and well aligned to the needs of policymakers and clinicians. Most examples and most evaluations of their effectiveness in supporting healthcare decision come from high-income countries (HIC).^{4,5} Experience with existing products (Box 1), and the required adaptations to formats for their effective use in low income countries, remain under researched. In this paper we report the findings of a mixed-method evaluation of the usefulness of these products, compared with more traditional formats, for a large, multidisciplinary guideline development panel in Kenya.

METHODS

Study design

This was a mixed-method study incorporating a randomised controlled trial (RCT) to assess the effectiveness of three different evidence summary formats with semi-structured follow-up interviews to explore panel members' views of these formats, experience with appraisal, use of and engagement with the research evidence.

Study setting

The Kenya Medical Research Institute - Wellcome Trust Research Programme together with Kenya's Ministry of Medical Services collaborated to host a one-week national guideline development workshop ('Child Health Evidence Week' [6]), held in Nairobi, Kenya, between 21st and 25th June, 2010.

Participants

Trial participants consisted of a multidisciplinary panel of healthcare professionals with varied roles in neonatal and paediatric policy and care in Kenya (table 1). All participants (n=70) had been nominated by their departmental heads to participate in the guideline development workshop. For the interviews, we purposively selected a sub-sample of stakeholders (n=16) who attended the guideline development workshop to: (1) achieve representation from groups involved at different levels of the health system; and (2) maximise the variety of views and experiences examined.

Interventions

We assembled evidence in three formats: systematic reviews (SR) alone (pack 'A'), systematic reviews with summary-of-findings tables (SR with SoF tables; pack 'B'), and 'graded-entry' formats (pack 'C'). Evidence pack 'A' represented the common standard practice of using systematic reviews and lengthy technical reports to inform healthcare policy and guideline development. Evidence pack 'B' represented the recently enriched format for preparing full Cochrane reviews.⁵

The 'graded entry' format (pack 'C') was designed to allow stepwise access to the evidence. It started with a 'front-end' short interpretation of the main findings and conclusions, drawn from evidence synthesis (webappendix 1). The 'front-ends' were adapted from existing review-derived formats designed to allow rapid scanning for key messages.⁷ Further development of these formats was guided by theoretical frameworks on information transfer for non-research audiences.^{8,9} Briefly, these frameworks propose that to improve knowledge transfer dissemination materials need to be easy to

use, accessible, credible (trustworthy), and desirable. The content needs to be current, relevant, methodologically competent, and comprehensive. Finally, the 'messages' need to be tailored to the specific needs and context of the users.

These front-end concise summaries were linked to a locally prepared, short, contextually framed, narrative report (hereafter called a narrative report)¹⁰ in which the results of the systematic review (and other evidence where relevant) were described and locally relevant factors that could influence the implementation of evidence-based guideline recommendations (e.g. resource capacity) were highlighted. The front-end summary and the narrative report were combined with the full systematic review (e.g. as published by the Cochrane Collaboration) to make a three-component set branded pack 'C'. The GRADE (Grading of Recommendations Assessment, Development, and Evaluation) system¹¹ was used to appraise and summarise evidence in summary-of-findings tables. These tables were included in the front-end summaries, narrative reports and were available to support stand-alone systematic reviews. The summaries were delivered to participants as pre-reading materials one month before the workshop.

Randomisation / Outcome measures

Evidence summaries in pack A, B and C formats were prepared for three 'tracer-interventions' relevant to neonatal care where new guidelines were being considered and for which systematic reviews had been recently published: (i) feeding regimens in sick newborns,¹² (ii) hand hygiene for infection prevention,¹³ and (iii) kangaroo care for low birth weight babies.¹⁴ For example, for feeding regimens we packaged evidence in the format of a systematic review (pack 'A'), a systematic review with summary-of-findings tables (pack 'B') and a 'graded-entry' format (pack 'C'); a similar approach was used to assemble differing evidence for hand hygiene and kangaroo care. We then provided all individual participants with evidence on all three tracer topics but used randomisation, within 5 professional

strata, to ensure that all participants received one tracer-topic with packaging approach A, one with packaging approach B, and one with packaging approach C. This enabled us, when examining which package is most effective at conveying research information, to address the possibility that the relationship is confounded by the nature of the topic and, in subsequent interviews, explore participants' views on all the 3 evidence packaging formats. Further details on the randomisation process is summarised in webappendix 2 and figure 1.

The primary outcome was the proportion of correct responses to key clinical questions relevant to the specific tracer topics. This tested the understanding of the effects of tracer-interventions on critical neonatal outcomes (mortality, morbidity). The secondary outcome measure was a composite score representing participants' self-reports of the clarity and accessibility of the evidence; participants rated their responses on a 3 or 5-point scale. These domains have successfully been tested and used in previous trials.^{5,15,16}

Data collection

Participants completed questionnaires on the first day (June 21st, 2010) of the guideline development workshop before the panel discussions about guidance recommendations. The questionnaires required responses from participants on: (1) personal characteristics (e.g. age, sex); (2) experience with interpreting research evidence and evidence based medicine skills; (3) recognition and understanding of key messages; (4) ease of access to and perceived value of provided evidence summaries; and (5) preferred summary formats. Participants were allowed up to 45 minutes to complete the questionnaire during which they had access to their personalized 'evidence packs'.

Individual face-to-face interviews were conducted (between July and August 2010) by NO following completion of the questionnaire but before analysis. Interview questions were derived from a review of

literature on best practices in clinical guideline development,¹⁷⁻¹⁹ pilot-tested (with three child health experts), and modified. The interviews lasted approximately 30 to 45 minutes and focused on collecting information about participants' experiences with appraisal, the use of research evidence and views on evidence summary formats.

Analysis

The sample size calculation was based on the pre-specified primary outcome assuming a two-way comparison of the three-component pack (C) versus the minimum standard pack (A), and a 1:1 allocation ratio. The trial was designed to have 90% power at an alpha (significance level) of 0.05 to detect a 40% absolute difference in the proportion of correct responses with 35 observations on each evidence pack (calculated using a conventional two-sided Chi-squared test on an intention-to-treat principle). Assuming a non-response rate of 10% we targeted at least 40 participants for randomisation. Analyses were done as per the original group allocations. All analyses treated the respondent as a unit of clustering.

The crude odds (likelihood) of correct responses for pack C compared to the odds for pack A (assumed baseline pack) were estimated using logistic regression. The models were also adjusted for the effect of strata and the tracer intervention. Pack by strata interaction terms were included to assess whether the effect of the different evidence packs differed by strata. The secondary outcome measure was the clarity and accessibility score. The mean clarity and accessibility scores of pack C compared to the mean scores of pack A were estimated using linear regression, and the lack of normality of the scores (figure 2) was accounted for using bootstrapping methods.²⁰ Similar processes were used to compare mean scores of pack B to A. To confirm the results of linear regression analysis, an alternative approach to the analysis was undertaken: the odds of a one point increase in the 'clarity and accessibility' scores of pack B and C compared to the odds of pack A were estimated using ordinal logistic regression models. Further

tests of interactions between pack and strata were performed using likelihood ratio tests. Where evidence of interaction was found, stratum-specific ORs and 95% CIs were calculated. All analyses were done with STATA (version 11.0). Further details on data handling and analysis is summarised in webappendix 3.

Audio interviews were transcribed verbatim by NO. Emerging themes and concepts were extracted by at least two co-investigators (NO, SS, NM, AF) working independently.²¹ These were compared and discussed; NO summarised the recurrent concepts into a set of initial descriptive themes with narrative summaries explaining each theme. These were discussed by investigators iteratively, with reference to the original interview transcripts, until a final set of themes was agreed upon.

Although quantitative and qualitative components of this study were pre-specified in the proposal integration occurred after separate analysis of each component at the point of interpretation of the results.

Ethical approval

Ethical approval for the conduct of this study was granted by the Kenya Medical Research Institute scientific committee and National Ethics Review Committee in Kenya (SSC Protocol No 1770).

Individual written informed consent for participation and audio recording of discussions was obtained prior to the face-to-face interviews. Confidentiality of participant information was ensured by assigning anonymous codes to individual audio interviews and transcripts.

RESULTS

The trial profile is summarised in figure 1. Seventy (91%) out of 77 participants invited attended the workshop. The most common reason for non-attendance was related to timing of the meeting. Table 1

shows the baseline characteristics of participants. A total of 16 participants were interviewed: three paediatricians from district hospitals, two senior paediatric trainees, one nurse manager, four trainers of healthcare workers (nursing and paediatric academics), one World Health Organization (WHO) officer, one research director, and four government policymakers.

Quantitative findings

Sixty-five (93%) of 70 participants completed questions on primary outcome measures. A descriptive summary of the proportion of correct and incorrect responses for each evidence pack is shown in table 2. There were no significant differences between packs in the odds of correct responses (adjusted ORs: pack B versus A 0.59, 95% CI 0.32 to 1.07; pack C versus A 0.66, 95% CI 0.36 to 1.21; $p=0.14$; table 3).

Assessment of the effects of summary formats across different groups of healthcare professionals suggested possible differences in the odds of correct responses across groups ($p=0.057$). Further subgroup analyses suggested that both pack B (systematic reviews with SoF tables) and pack C (the three component, graded entry pack) improved understanding for policymakers (pack B: OR 1.5, 95% CI 0.15 to 15.15; pack C: OR 1.5, 95% CI 0.64 to 3.54) and trainee paediatricians (pack B: OR 1.3, 95% CI 0.37 to 4.66; pack C: OR 1.78, 95% CI 0.43 to 7.33).

'Graded-entry' formats (pack C), compared to systematic review formats (pack A), were associated with a significantly higher mean 'value and accessibility' score (adjusted mean difference 0.52, 95% CI: 0.06 to 0.99; table 3). Similarly, pack C, compared to pack A, was associated with a 1.5 higher odds of judgments about the quality of evidence for critical neonatal outcomes being clear and accessible (adjusted OR: 1.52, 95% CI 1.06 to 2.20; table 3). There was no evidence that pack B (systematic reviews with SoF tables) improved this composite score compared to pack A (adjusted mean difference: 0.11, 95% CI - 0.71 to 0.48; table 3).

More than half of the respondents (60%) found systematic reviews to be more difficult to read compared to narrative reports, but some (17%) responded that systematic reviews were easy to read. About half of the participants (51%) found systematic reviews to be easier to read compared to summary-of-findings tables (26%). A higher proportion of participants preferred evidence summarised in narrative report formats to the full version of the systematic reviews (53% versus 25%). See table 4 for a complete overview of these findings.

Participants' self-rated experiences with research literature and familiarity with evidence based medicine terminology are presented in table 5. Of note are that about 29% of respondents had not read a systematic review in the last year. Most of the participants (53[82%]) responded that they spent less than 12 hours reading provided evidence summaries before the workshop and 6 (9%) reported not reading materials at all.

Qualitative findings

Views on the different formats for presenting systematic review evidence (see panel 1 for illustrative quotes).

The majority of participants interviewed found research information summarised in the form of narrative reports to be clearer, easy to read, easy to understand, and containing 'just the right amount of information'. Conversely, participants expressed considerable variability in views for systematic reviews and GRADE summary-of-findings tables: while some found the comprehensive and structured nature of information presentation in systematic reviews to be useful, a number expressed difficulties with extracting pertinent information. Some participants found GRADE summary tables to be good for 'rapid consultation'; however, a number of participants found them difficult to understand as stand-alone summaries. For example, one participant expressed difficulties with distinguishing between the

GRADE categories of quality of evidence. Of note, many participants reported lack of time and the volume of evidence as factors contributing to more complete engagement with the evidence.

Panel 1. Panelists' views on the research summary and presentation formats

[006]: 'The mini-review [narrative report] I think was like one of the most useful things I came across, because I found that they had just the right amount of information...and also the Cochrane reviews I found were very good'

[010]: '...this particular one [table-of-key findings] was not very easy for me to understand...sometimes making a clear cut difference between especially the quality grades was not very easy...alone on itself I found it difficult...I needed to go back to the summaries, read a bit and come and compare...one needed to have some prior knowledge especially on how to interpret it...'

[012]: 'The systematic reviews give more information on what was done in the research but you will need a lot of time to go through it. So if I were to be asked which one I preferred, I would pick the mini-review [narrative report] because it was very brief and easy to understand. It's easy to understand and you don't go and go until you get tired along the way'

[014]: 'The Cochrane is good because it's comprehensive and a very good summary of the work that's available as well as what has been done. Its main disadvantage is that it takes more time to read through and digest. The mini-reviews [narrative reports] were the most useful because they were a condensed version, so they are easy to use even when you have limited time they will prove useful. The table-of-key findings is good for a rapid consultation but doesn't give you a feel of how those conclusions were arrived at'

[016]: 'The problem with the reviews is that most of them had limited data...But I was feeling that a lot

more other data probably from observation studies could be available and would have been of more value. But at the end of the day you base most of your decisions on randomised controlled trials, but you should know a little more of what was done in the other studies'

Pre-workshop materials

[006]: '...like the people I was sitting with, most of them had not gone through all of the material, a few had gone through about half and a few had not looked at it at all'

[008]: 'No, the reading of the materials appeared not to have been exhausting accept you backed it up by making presentations...'

[010]: 'Generally I think very few of the participants went through the materials before coming to the workshop. Personally I would say I didn't spend adequate time that would have enabled me to understand the material better, for better participation during the workshop...'

[012]: '...and the volume of the document was another challenge. With the limited time people had and the volume of the document, people had very little time to go through them'

[015]: 'The people I interacted with that Monday morning...most of them said they were not able to read from page to page...mostly the ones I interacted with were saying that the only time they could read is in their homes, and they were saying they didn't have time when they reach home; there is a lot of activities...'

Experiences with appraisal and use of research evidence (panel 2). The majority of participants responded that they were not conversant with assessing the quality of scientific literature or evidence-based medicine terminology, such as PICO (Patient, Intervention, Comparison, Outcome). A number

suggested that a short course on evidence-based medicine would be beneficial to support evidence-based guideline development.

Panel 2. Panelists' experiences with appraisal and use of research evidence

[004]: 'No, I can't say I was comfortable, even when the PICO [Patient, Intervention, Comparison, Outcome] format was introduced I personally had problems understanding it...but now I understand it very properly and even the GRADE system, for me it really sunk well after the evidence week'

[011]: 'Not at all, even being able to synthesize a research paper...it's still a process that I would like to learn. For many of the participants I think a short course in that area would be useful and it would help a bit also in the voting'

[013]: 'Most of us take it [research information] as it's given to us; we don't sieve it and see whether it's of high or low quality...the general practitioners don't really know the difference of [between] these studies'

[015]: '...I didn't know there is low quality, moderate [quality] or even when I find in the books that it is low quality, I did not know actually what they based it on'

DISCUSSION

Interpretation of results

The current study assessed the effectiveness and panelists' views and experiences with different evidence summary formats. The quantitative evaluation showed that 'graded-entry' evidence summary formats which included narrative reviews highlighting local factors relevant for implementation,

compared to full systematic reviews, were associated with higher composite scores for clarity and accessibility of information on critical outcomes. However, the proportion of correct answers to questions testing the ability to identify, access and reproduce key messages from the evidence summaries was not improved. This failure may reflect a true lack of effect of format on this outcome. Alternatively, it may be due to inadequate power to detect a smaller but useful effect or inadequate participant preparation in those with limited prior exposure to research evidence (panel 1 and table 5). Although not statistically significant, results of stratified analyses provide some support for the latter possibility as performance of policymakers and trainee paediatricians, arguably more familiar with interpreting evidence, was somewhat better. This interpretation of the trial-findings seen together with the qualitative data showing difficulties engaging with evidence among some panelists (panel 2), suggests that improving basic skills in evidence-based medicine may be important as we seek to engage wider, non-expert audiences in guideline development approaches. Strategies to enhance panelists' evidence retrieval and synthesis skills may include: (1) involvement of panel members in the preparatory stages of guideline development (e.g. in the conduct of systematic reviews)^{22,23}; and (2) education of panel members on guideline development methods (e.g. on the GRADE system).

Despite an absence of effect on our main outcome, 'graded-entry' summary formats and their components were reported to: improve clarity of information presentation, improve accessibility to key information, be more reader-friendly (both narrative reports and summary-of-findings tables), and were preferred by participants over systematic reviews. These favorable findings on secondary outcomes are further supported by qualitative findings (panel 1). A preference for narrative reports may be due to: (1) their abbreviated and plain language nature; (2) incorporation of judgments on the quality of evidence for guideline relevant outcomes; and (3) inclusion of contextual information (e.g. local antimicrobial resistance patterns). A comparison of our findings with related studies is summarised in Box 2. However,

it is clear that such 'graded-entry' formats are not by themselves an answer to the problem of improving the use of evidence in multidisciplinary guideline panels comprising important policy and provider groups - something that needs to be taken into account when structuring future approaches.

Box 2. Research in context

To compare current results with related studies we searched the Cochrane Library and PubMed (both from database inceptions up to March 2012) for articles examining ways of presenting evidence for use in healthcare decision-making (full search strategy available from the authors on request). The search identified six related studies: three systematic reviews,^{3,24,25} two randomised controlled trials,^{5,26} and one qualitative study.¹⁶ Only one²⁶ of the studies was conducted in the context of clinical guideline development. The systematic reviews noted that presentation of results of health technology assessments²⁴ and systematic reviews^{3,25} using 'graded-entry' formats (e.g. one page take home messages, a three-page executive summary and a 25-page report, 1:3:25 format) rendered them more useful to healthcare managers and policymakers. Research syntheses using 'graded-entry' formats were however reported to be rare (identified in only 7% of 45 websites searched).³

In the first randomised controlled trial¹⁵ participants attending an evidence-based practice workshop (N=72; Norway) and a Cochrane Collaboration entities meeting (N=33; UK) were allocated to receive a Cochrane review with or without a summary-of-findings table. Inclusion of a summary-of-findings table was found to significantly improve understanding and rapid retrieval of key findings. In the second randomised controlled trial,²⁶ paediatricians and trainee paediatricians in private and public practice in Mexico (N=216) were allocated to receive the same clinical recommendation presented using four different grading systems: NICE (National Institute for Health and Clinical Excellence), SIGN (Scottish Intercollegiate Guidelines Network), GRADE and Oxford CEBM (Centre for Evidence-Based Medicine).

The primary outcome was mean change (before compared to after reading the guideline) in simulated clinician decision to use a therapy. The findings showed that clinician's decision to use a therapy was influenced most by the GRADE system. However, no significant differences were found between systems in the clarity of presentation of care recommendation (defined as 'ease of reading and understanding').

The qualitative interview (N=13) study¹⁶ explored health policymakers' and managers' experiences with short summaries of systematic reviews in five low-and middle-income countries, LMICs (Argentina, China, Colombia, South Africa, Uganda). The findings showed that participants preferred 'graded-entry' formats, particularly the section on the 'relevance of the summaries for LMICs, which compensated for the lack of locally-relevant detail in the original review'.

Overall, observed differences in results across systematic reviews and primary studies examining impacts of evidence presentation formats, including our own, could be explained by differences in: study designs (randomised controlled trials versus user surveys), presentation formats (due to variations in layouts, contents, visual aids), types of participants (multidisciplinary versus unidisciplinary; variations in evidence based behaviors and skills); study contexts (evidence based medicine training versus policymaking environments); definitions of outcome measures (e.g. in Garcia et al,²⁶ unlike in other studies, clarity of presentation of recommendation was defined as 'ease of reading and understanding'); and durations of exposure to the evidence summaries. These differences underscore the need for a unified system for evaluating impacts of evidence summary and presentation formats to enhance comparability and interpretation of results of accumulating work in this field.

Implications for development of health systems guidance in LMICs

We believe the results of our evaluation provide valuable inputs into methods for summarising and communicating evidence to those charged with formulating recommendations in LMICs. First, there is a need for wider sensitization to the principles of and tools for evidence-informed decision making. We also found our narrative reports, that adapted the GRADE system to synthesize and present evidence on treatment effects and on-the ground (contextual) factors and values that might influence intervention effectiveness, to be liked by busy decision-makers. Such narrative reports could be made even more policy-relevant by inclusion of information on wider health systems perspectives such as: potential structural changes needed for effective implementation, equity, feasibility of scaling up and monitoring and evaluation issues.

Strengths and limitations

Combining a quantitative RCT design with qualitative interviews and using both in our interpretation of data improved our understanding of, and confidence in the study results. Our decision to study multiple tracer-topics rather than focus on a single guideline development topic may be considered a weakness or a strength. Perhaps if participants had only to read the literature provided for one topic, 'graded-entry' formats may have proved superior even on the primary outcome. However, in a low-income country setting like ours, it seems unlikely that countries will have the resources in the near future to engage specific panels for meetings for each guideline topic. Thus, we deliberately wished to examine a potentially efficient approach in which the panel considered multiple topics.

Future research

Although our findings suggest possible benefits of 'graded-entry' formats, evidence about the best formats for improving actual understanding and use of research findings in decision making processes (a key aim in knowledge translation) appears to be lacking. Thus, further studies are needed to learn more

about the relative effectiveness of the various evidence summary formats that are available, and how these may be combined with alternative approaches.

CONCLUSIONS

‘Graded-entry’ formats were found to improve clarity and accessibility of research evidence, although this was not reflected in the number of correct responses to key clinical questions. Providing a ‘front-end’ summary of key information linked to locally relevant factors that support implementation, and the full systematic review, may help those developing guidelines access and contextualise research evidence.

Acknowledgement

We are grateful to Judy Nganga for her help with the organisation of the workshop. We thank all participants for their time and contributions during the guideline development workshop. This work is published with the permission of the Director of KEMRI.

Author contributions

Study design: NO, SS, EA, AF, RN, ME. Data collection: NO, SS, NM, AF, RN, ME. Data analysis: EA, NO, SS, AF, NM, ME. Manuscript preparation: NO prepared the first draft; all authors reviewed and approved the final manuscript. All authors had full access to the data derived from the study and can take full responsibility for the integrity of the data and the accuracy of the data analyses.

Funding

NO and NM were supported by funds from a Wellcome Trust Strategic Award (#084538). ME is supported by a Wellcome Trust Senior Fellowship (#076827). The funding source had no role in study design, data collection, analysis, data interpretation, writing of the report or the decision to submit for publication. All authors carried out this research independently of the funding body.

Competing interests

None declared.

REFERENCES

1. Mulrow CD. Rationale for systematic reviews. *BMJ* 1994;309(6954).
2. Tugwell P, Robinson V, Grimshaw J, Santesso N. Systematic reviews and knowledge translation. *Bull World Health Organ* 2006;84:643-651.
3. Lavis J, Davis H, Oxman A, Denis JL, Golden-Biddle K, Ferlie E. Towards systematic reviews that inform health care management and policy-making. *J Health Serv Res Policy* 2005;10:35-48.
4. GRADE working group. <http://www.gradeworkinggroup.org/> [accessed 4 April, 2012].
5. Rosenbaum SE, Glenton C, Oxman AD. Summary-of-findings tables in Cochrane reviews improved understanding and rapid retrieval of key information. *J Clin Epidemiol* 2010;63:620-626.
6. Opiyo N, Shepperd S, Musila N, English M, Fretheim A. Developing clinical practice guidelines for care of childhood illnesses in Kenya: a descriptive study. *J Clin Epidemiol*. In-press.
7. Communication notes. Reader-friendly writing - 1:3:25. Canadian Health Services Research Foundation; 2001. http://www.chsrf.ca/Migrated/PDF/CommunicationNotes/cn-1325_e.pdf [accessed 4 April, 2012].
8. Dobbins M, Rosenbaum P, Plews N, Law M, Fysh A. Information transfer: what do decision makers want and need from researchers? *Implement Sci* 2007;2.
9. Morville P. User experience design. <http://semanticstudios.com/publications/semantics/000029.php> [accessed 4 April, 2012].
10. Improving delivery of hospital care in Africa (iDOC Africa). Available at: <http://www.idoc-africa.org/etat-evidence-base/child-health-evidence-week-chew-2010.html> [accessed 4 April, 2012].

11. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924-926.
12. Edmond K, Bahl R. Optimal feeding of low-birth-weight infants. Technical review. World Health Organisation, 2006.
13. World Health Organisation (WHO). WHO guidelines on hand hygiene in health care. Geneva: World Health Organisation, 2009.
14. Conde-Agudelo A, Diaz-Rossello JL, Belizan JM. Kangaroo mother care to reduce morbidity and mortality in low birthweight infants. *Cochrane Database Syst Rev* 2003;CD002771.
15. Rosenbaum SE, Glenton C, Glenton C, Cracknell J. User experiences of evidence-based online resources for health professionals: user testing of The Cochrane Library. *BMC Med Inform Decis Mak* 2008;8:34.
16. Rosenbaum SE, Glenton C, Wiysonge CS, Abalos E, Mignini L, Young T, et al. Evidence summaries tailored to health policy-makers in low- and middle-income countries. *Bull World Health Organ* 2011;89:54-61.
17. Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, Askham J, et al. Consensus development methods, and their use in clinical guideline development. *Health Technol Assess* 1998; 2(3): i-iv, 1-88.
18. National Institute for Health and Clinical Excellence (January 2009). The guidelines manual. London: National Institute for Health and Clinical Excellence. Available from: www.nice.org.uk
19. World Health Organisation (WHO). WHO handbook for guideline development. WHO 2008.
20. Henderson AR. The bootstrap: a technique for data-driven statistics. Using computer-intensive analyses to explore experimental data. *Clin Chim Acta* 2005;359(1-2):1-26.

21. Pope C, Ziebland S, Mays N. Qualitative research in health care. Analysing qualitative data. *BMJ* 2000; 320(7227):114-116.
22. English M, Opiyo N. Getting to grips with GRADE-perspective from a low-income setting. *J Clin Epidemiol* 2011;64(7):708-710.
23. Duke T, Campbell H, Ayieko P, Opiyo N, English M, Kelly J, et al. Accessing and understanding the evidence. *Bull World Health Organ* 2006;84(12):922.
24. Lavis JN, Wilson MG, Grimshaw JM, Haynes RB, Ouimet M, et al. Supporting the use of health technology assessments in policy making about health systems. *Int J Technol Assess Health Care* 2010;26:405-414.
25. Chambers D, Wilson PM, Thomson CA, Hanbury A, Farley K, Light K. Maximizing the impact of systematic reviews in health care decision making: a systematic scoping review of knowledge-translation resources. *Milbank Q* 2011;89:131-156.
26. García CA, Alvarado KP, Gaxiola GP. Grading recommendations in clinical practice guidelines: randomised experimental evaluation of four different systems. *Arch Dis Child* 2011;96:723-728.

Tables and Figures

Table 1. Profile of guideline panel members

Characteristic	Frequency	%
Sex		
Male	34	49
Female	36	51
Age (years)		
21-30	8	11
31-40	35	50
41-50	14	20
51-60	10	14
Above 60	3	4
Profession		
Paediatrician	32	46
Medical Officer	11	16
Nursing Officer	7	10
Research (research supervision) role	5	7
Trainer of healthcare workers	5	7
National or provincial role for MoM / MoPHS	4	6
Clinical Officer	3	4
Pharmacist	2	3
Academic administration	1	1
Number of years as a healthcare professional		
3-7	23	33
8-12	14	20
13-21	16	23
22-40	17	24

Table 2. Proportion of incorrect and correct responses (N=65 participants)

Responses	Pack A	Pack B	Pack C	Total
Incorrect	39 (30.0)	51 (39.2)	49 (38.0)	139 (35.7)
Correct	91 (70.0)	79 (60.8)	80 (62.0)	250 (62.3)
Total	130	130	129	389

Figures in brackets are percentages; pack A=systematic review; pack B=systematic review with summary-of-findings table; pack C='graded-entry' format (a 'front-end' summary of key information linked to a short contextually framed narrative report and full systematic review)

Table 3. Primary and secondary outcomes (N=65 participants)

Odds ratios for correct responses				
Pack	Unadjusted OR (95% CI)	p-value	Adjusted OR (95% CI)†	p-value
Pack A	1 (Referent)	-	1	-
Pack B	0.66 (0.37-1.20)	0.366	0.59 (0.32-1.07)	0.220
Pack C	0.70 (0.39-1.27)		0.66 (0.36-1.21)	
Mean differences in 'value and accessibility' scores				
	Mean difference (95% CI)	p-value	Mean difference (95% CI)†	p-value
Pack A	1	-	1	-
Pack B	-0.14 (-0.77 to 0.49)	0.046	-0.11 (-0.71 to 0.48)	0.025
Pack C	0.49 (0.01 to 0.98)		0.52 (0.06 to 0.99)	
Odds ratios for 'value and accessibility' scores				
	Unadjusted OR (95% CI)	p-value	Adjusted OR (95% CI)†	p-value
Pack A	1	-	1	-
Pack B	0.97 (0.64-1.47)	0.038	0.91 (0.57-1.46)	0.022
Pack C	1.48 (1.06-2.08)		1.52 (1.06-2.20)	

OR=odds ratio; CI=confidence interval; † Adjusted for type of strata and tracer-intervention; pack A= systematic review; pack B= systematic review with summary-of-findings table; pack C='graded-entry' format (a 'front-end' summary of key information linked to a short contextually framed narrative report and full systematic review)

Table 4. Ease of use and summary format preferences

	n/N [†]		
	Percentage (95% confidence interval)		
	Easy to read	Neutral	Difficult
How easy to read did you find evidence summarised in systematic review formats compared to narrative report formats	10/58 17% (8%-27%)	13/58 22% (12%-33%)	35/58 60% (48%-73%)
How easy to read did you find evidence summarised in systematic review formats compared to summary-of-findings table formats	16/61 26% (15%-37%)	14/61 23% (13%-33%)	31/61 51% (38%-63%)
	Strongly prefer systematic review	Neutral	Strongly prefer narrative report
Comparing systematic review versus narrative report formats	16/64 25% (14%-36%)	14/64 22% (12%-32%)	34/64 53% (41%-65%)
	Strongly prefer systematic review	Neutral	Strongly prefer SoF table
Comparing systematic review versus summary-of-findings (SoF) table	20/63 32% (20%-43%)	18/63 29% (17%-40%)	25/63 40% (27%-52%)
	Strongly prefer narrative report	Neutral	Strongly prefer SoF table
Comparing narrative report versus summary-of-findings (SoF) table	24/64 38% (25%-50%)	25/64 39% (27%-51%)	15/64 23% (13%-34%)

[†] Denominators exclude missing data

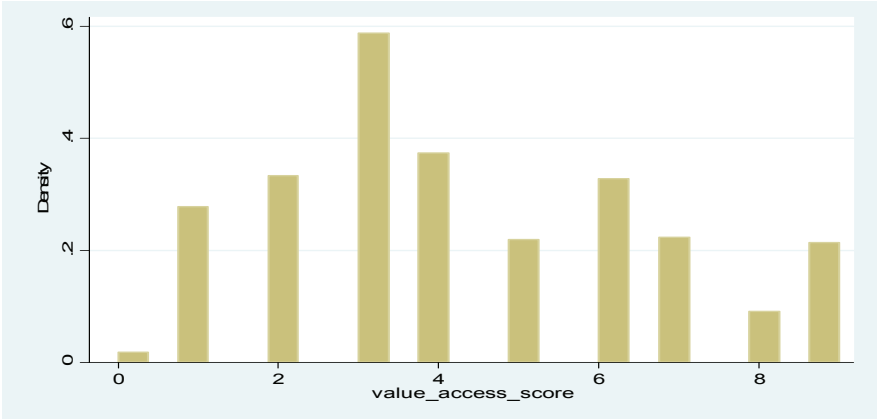
Table 5. Participants' self-rated experience with research literature and familiarity with evidence-based medicine terminologies†

Characteristic	Frequency (%)
Frequency of reading journal articles in the last one year (n=65)	
Less than once per year	16 (24.6%)
1 to 4 times per year	15 (23.1%)
5 to 10 times per year	13 (20.0%)
More than 10 times per year	21 (32.3%)
Frequency of reading systematic review in the last 1 year (n=65)	
Never before now	19 (29.2%)
Less than once per year	14 (21.5%)
1 to 4 times per year	21 (32.3%)
More than 5 times per year	11 (16.9%)
Time spent reading provided evidence summaries (n=65)	
< 12 hours	53 (81.6%)
> 12 hours	6 (9.2%)
Did not read	6 (9.2%)
Confidence in interpreting the term 'randomisation' (n=65)*	
Very confident	20 (30.8%)
Confident	33 (50.8%)
Not so confident	12 (18.5%)
Description of a systematic review (n=63)	
Correct	39 (61.9%)
Incorrect	24 (38.1%)
Interpretation of risk ratio (n=64)	
Correct	32 (50.0%)
Incorrect	32 (50.0%)
Calculation of risk ratio of mortality comparing two treatments (n=65)	
Correct	16 (24.6%)
Incorrect	49 (75.4%)

† Denominator exclude missing data

*Similar results were obtained for responses to questions 'confidence in interpreting the term blinding' and 'confidence in interpreting the term selection bias'

Figure 2. Distribution of ‘value and accessibility’ scores



Box 1. Examples of research synthesis (systematic review derived) products

Type of synthesis	Target audience	Description (Scope)	Source
Abstracts of Reviews of Effects	Policymakers, clinicians	Quality-appraised summary of single non-Cochrane systematic reviews of effects of healthcare or health system interventions (delivery and organisation of health services)	Centre for Reviews and Dissemination (CRD) http://www.crd.york.ac.uk/CMS2Web/AboutDare.asp
Cochrane PICO	Healthcare practitioners and decision-makers	Short summaries (about one page) of a clinical question addressed by one or more Cochrane reviews	Cochrane Editorial Unit http://editorial-unit.cochrane.org/cochrane-pico
Evidence Update	Healthcare practitioners, policymakers	Two-page summaries of a Cochrane review of effects of healthcare or health system interventions relevant to people in low and middle-income countries (LMICs)	Effective Health Care Research Programme Consortium http://www.liv.ac.uk/evidence/evidence.htm
Plain language summaries of systematic reviews	Policymakers, researchers, clinicians	Structured summaries of systematic reviews of effects of healthcare or health systems interventions in maternal and child health in LMICs Summaries prepared using GRADE system	SUPPORT Collaboration http://www.support-collaboration.org/summaries.htm
Policy briefs	Policymakers, researchers, civil societies	Summaries of policy-relevant reviews (healthcare, health systems, behaviour-change interventions) with a focus on LMICs Summaries prepared using 'graded entry' formats (i.e., a list of key messages, an executive summary,	SURE http://www.evipnet.org/local/SURE%20Website/home%20page.htm

	and a full report)	EVIPNet http://www.evipnet.org/local/SURE%20Website/home%20page.htm
Policy Liaison Initiative summaries	Policymakers	Summaries of policy-relevant Cochrane reviews of effects of healthcare, health system, behaviour change and consumer-targeted interventions http://www.cochrane.org.au/ebpnetwork/
Structured evidence summaries	Healthcare professionals, and consumers	Summaries of multiple primary studies, systematic reviews, or evidence-based clinical guidelines of diagnostic, therapeutic and delivery of healthcare interventions http://connect.ibiconnectplus.org/
Synopses of reviews and single articles	Clinicians	Structured abstracts of systematic reviews and original studies of diagnostic and therapeutic interventions The American College of Physicians (ACP) Journal Club http://acpic.acponline.org/shared/burpouse_and_procedure.htm
2-page summary statements	Healthcare practitioners, decision-makers, and managers	2-page summary statements synthesizing results of systematic reviews of health promotion and public healthcare interventions Health-evidence.ca http://www.health-evidence.ca/html/AboutUs

GRADE=Grading of Recommendations Assessment, Development and Evaluation; SUPPORT=Supporting Policy Relevant Reviews and Trials; SURE=Supporting the Use of Research Evidence; EVIPNet=Evidence-Informed Policy Network; PICO=Population (Patients, Problem), Interventions (Exposure), Comparison (Control), Outcomes

Webappendix 1. 'Front-end' evidence summary

Evidence summary: Kangaroo mother care for low birth weight infants

Clinical need

Kangaroo mother care (KMC), defined as continuous skin to skin between a mother and her newborn allowing frequent and exclusive breastfeeding, has been proposed as an alternative to conventional care (incubators, cots, etc) for low birth weight (LBW, less than 2500 g) babies. The evidence for the effectiveness and safety of KMC versus conventional care in LBW infants is considered in this summary.

Clinical questions, Quality of evidence† and Key findings

- **Population:** Low birth weight infants, less than 2500 g
- **Comparisons:** Kangaroo mother care versus conventional care
- **Outcomes:** Neonatal mortality, morbidity, breastfeeding status, costs and length of hospital stay

1. What is the evidence that KMC reduces the risk of mortality in LBW infants?

Key findings

- **Low quality evidence** suggests that KMC does not reduce the risk of death in stabilized LBW infants
- **Low quality evidence** suggests that KMC may reduce the risk of death in LBW infants if initiated very early in life before stabilization

2. What is the evidence that KMC reduces morbidity in LBW infants?

Key findings

- **Low quality evidence** suggests that KMC reduces the risk of morbidity (mild / illnesses, nosocomial infections) in LBW infants

3. What is the evidence that KMC improves breastfeeding outcomes in LBW infants?

Key findings

- **Low quality evidence** suggests that KMC increases the likelihood of exclusive breastfeeding at discharge in LBW infants
- **Moderate quality evidence** suggests that KMC increases the likelihood of exclusive breastfeeding at 41 weeks corrected age in LBW infants
- **Very low quality evidence** suggests that KMC may improve the chances of exclusive breastfeeding of LBW infants at the age of six months post birth

4. What is the evidence that KMC reduces the length of hospital stay of LBW infants?
<p>Key findings</p> <ul style="list-style-type: none"> • Low quality evidence suggests that LBW babies on KMC stay hospitalized for a shorter duration compared to those on conventional care
5. What is the evidence for the cost-benefit of KMC compared to standard neonatal care?
<p>Key findings</p> <ul style="list-style-type: none"> • Very low quality evidence suggests that the cost of care for babies on KMC is lower than the costs of standard care

‡ Quality of evidence is categorized as 'high', 'moderate', 'low' or 'very low'.

- **HIGH:** Further research is very unlikely to change our confidence in the estimate of effect
- **MODERATE:** Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
- **LOW:** Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
- **VERY LOW:** We are very uncertain about the estimate

Quality of Evidence and Summary of Findings

Question 1: What is the evidence that KMC reduces the mortality risk in LBW infants?

Intervention: Kangaroo mother care

Comparison: Conventional care

Bibliography: Worku et al³; Charpak et al⁴; Sloan et al⁵; Cattaneo et al⁵; Suman et al⁹

Quality assessment							Summary of findings		Importance
No of studies	No of infants	Design	Limitations	Inconsistency	Indirectness	Imprecision	Effect size (95% CI)	Quality (GRADE)	
Mortality before stabilization (follow-up 4 to 6 days)									
1	123	randomised controlled trial	serious†	no serious inconsistency	no serious indirectness	serious‡	RR 0.57 (0.33 to 1.0)	⊕○○○ LOW	CRITICAL
Mortality after stabilization (follow-up 1 to 7 weeks)									
4	1512	randomised controlled trials	serious††	no serious inconsistency	no serious indirectness	no serious imprecision	RR 0.70 (0.41 to 1.21)	⊕⊕⊕○ MODERATE	CRITICAL

†significant number of recruited infants not randomised, blinding of investigators / data collectors unclear; ‡ small sample size, wide 95% confidence interval; †† blinding of investigators / data collectors unclear, potential for reporting bias (selective reporting of outcomes)

1	275	randomised controlled trial	serious¶	no serious inconsistency	no serious indirectness	serious‡	RR 0.90 (0.3 to 0.66)	⊕⊕○○ LOW	CRITICAL
Morbidity (risk of severe illness at 1 year corrected age; follow-up 0 to 12 months)									
1	285	randomised controlled trial	serious††	no serious inconsistency	no serious indirectness	serious‡	RR 0.95 (0.06 to 15.09)	⊕⊕○○ LOW	CRITICAL

‡blinding of intervention to both investigators and data collectors unclear; ‡few number of events; ¶unclear concealment of intervention allocation, unblinded outcome assessment; ††significant loss to follow-up

2	759	randomised controlled trial	serious [†]	no serious inconsistency	serious [‡]	serious ^{††}	RR 0.92 (0.69 to 1.23)	⊕○○○ VERY LOW	IMPORTANT
---	-----	-----------------------------------	----------------------	-----------------------------	----------------------	-----------------------	---------------------------	-------------------------	-----------

[†]undclear concealment of allocation of interventions / blinding of outcome assessment; [‡]one of the included studies conducted in a high income setting with a lactation consultant; [¶]undclear concealment of allocation of interventions; [†]small number of events (<300)

Question 4: What is the evidence that KMC reduces the length of hospital stay of LBW infants?

Intervention: Kangaroo mother care

Comparison: Conventional care

Bibliography: Kadam et al⁸; Gathwala et al¹⁰; Boo et al¹³

Quality assessment							Summary of findings		Importance
No of studies	No of infants	Design	Limitations	Inconsistency	Indirectness	Imprecision	Effect size (95% CI)	Quality (GRADE)	
Length of hospital stay (follow-up 0 to 40 days)									
1	126	randomised controlled trial	serious†	no serious inconsistency	no serious indirectness	serious‡	13.5 days (KMC) versus 22.5 days (CMC)	⊕⊕○○ LOW	IMPORTANT
Length of hospital stay (follow-up 0 to 2 months)									
2	199	randomised controlled trial	serious†	serious¶	no serious indirectness	serious‡	KMC = shorter duration of hospital stay††	⊕○○○ VERY LOW	IMPORTANT

†unclear concealment of allocation of interventions / blinding of outcome assessment /selective reporting of outcomes; ‡small number of enrolled participants; CMC - conventional method of care; ¶age at randomisation into KMC group inconsistent across included studies;

††Gathwala et al; (KMC, 3.56 days versus CMC, 6.8 days), Kadam et al (KMC, 8.5 days versus CMC, 9.3 days)

Question 4: What is the evidence that KMC reduces the length of hospital stay of LBW infants?

Intervention: Kangaroo mother care

Comparison: Conventional care

Bibliography: Sloan et al⁵; Cattaneo et al⁶

Quality assessment							Summary of findings		Importance
No of studies	No of infants	Design	Limitations	Inconsistency	Indirectness	Imprecision	Effect size (95% CI)	Quality (GRADE)	
Cost of care (follow-up 0 to 6 months)									
2	560	observational studies	very serious†	no serious inconsistency	serious‡	no serious imprecision	KMC = Lower costs of care¶	⊕○○○ VERY LOW	IMPORTANT

[†]potential for selection / investigator bias in recruitment of participants / measurement of outcomes; [‡]items costed unclear; [¶] Sloan et al (KMC, US \$101 versus CMC, US \$130), Cattaneo et al (KMC, US \$7,501 versus CMC, US \$9,876)

Characteristics of the evidence

This evidence summary is based on a comprehensive search and critical appraisal (for methodological rigor and clinical practice applicability) of best currently available literature. The evidence in this summary comes from:

- One Cochrane review of randomised controlled trials (RCTs) (N=1,362 infants, 3 studies)¹
- One overview of 2 systematic reviews and 7 RCTs²
- Ten RCTs (N=2,086 infants)³⁻¹²

References

1. Conde-Agudelo A, Belizán J. Kangaroo mother care to reduce morbidity and mortality in low birthweight infants. *Cochrane Database Syst Rev* 2003, Issue 2. Art. No.: CD002771. DOI: 10.1002/14651858.CD002771.
2. Blackwell K, Adriano C. What is the evidence of Kangaroo mother care of the very low birth weight infants. International Child Health Review Collaboration (ICHRC). Available at: www.ichrc.org/pdf/kangaroo.pdf (Accessed May 17, 2010).
3. Worku, B, Kassie A. Kangaroo Mother Care: A randomized controlled trial on effectiveness of early kangaroo mother care for the low birth weight infants in Addis Ababa, Ethiopia. *J Trop Ped* 2005; 51(2); 93-7.
4. Charpak N, Ruiz-Peláez JG, Figueroa de C Z, Charpak Y. A randomized, controlled trial of kangaroo mother care: results of follow-up at 1 year of corrected age. *Pediatrics* 2001;108(5); 1072-1079.
5. Sloan NL, Camacho LW, Rojas EP, Stern C. Kangaroo mother method: randomised controlled trial of an alternative method of care for stabilized low-birthweight infants. Maternidad Isidro Ayora Study Team. *Lancet* 1994; 344(8925); 782-785.
6. Cattaneo A, Davanzo R, Worku B, Surjono A, Echeverria M, Bedri A, et al. Kangaroo mother care for low birth weight infants: a randomised controlled trial in different settings. *Acta Paed* 1998; 87; 976-985.
7. Charpak N, Ruiz-Peláez JG, Figueroa de C Z, Charpak Y. Kangaroo mother versus traditional care for newborn infants ≤ 2000 grams: a randomised controlled trial. *Pediatrics* 1997;100(4):682-688.
8. Kadam S, Binoy S, Kanbur W, Mondkar JA, Fernandez A. Feasibility of kangaroo mother care in Mumbai. *Indian J Ped* 2005;72; 35-38.
9. Suman RP, Udani R, Nanavati R. Kangaroo mother care for low birth weight infants: a randomized controlled trial. *Indian Pediatr* 2008; 45; 17-23.
10. Gathwala G, Singh B, Balhara B. KMC facilitates mother baby attachment in low birth weight infants. *Indian J Ped* 2008; 75; 43-47.
11. Hake-Brooks SJ, Anderson S. Kangaroo mother care and breast-feeding of mother-preterm infant dyads 0-18 months: A randomized, controlled trial. *Neonatal Network* 2008; 27; 151-159.
12. Rojas MA, Kaplan M, Quevedo M, Sherwonit E, Foster LB, Ehrenkranz RA, et al. Somatic growth of preterm infants during skin-to-skin care versus traditional holding: A randomized, controlled trial. *Developmental and Behavioral Pediatrics* 2003; 24; 163-168.

Webappendix 2. Randomisation

Evidence summaries in pack A, B and C formats were prepared for three 'tracer-interventions' relevant to neonatal care where new guidelines were being considered and for which systematic reviews had been recently published: feeding regimens in sick newborns (#1) [12], hand hygiene for infection prevention (#2)[13], and kangaroo care for low birth weight babies (#3)[14].

We defined the following 'evidence pack + tracer-intervention' combination sets. Each participant received as their pre-reading material one of these three 'evidence pack + tracer-intervention' combination sets.

- A#1, B#2, C#3
- A#2, B#3, C#1
- A#3, B#1, C#2

With the three evidence packs, and three 'tracer-interventions' a complete experimental design would have resulted in six possible combinations. However, we used a simplified and partial experimental design for reasons of feasibility.

We expected experience with use of research evidence to vary across the stakeholder groups invited to the guideline development workshop. We thus grouped participants into the following 5 strata according to their roles: non-specialist health workers (nurses and clinical officers) (stratum 1); expert clinicians (paediatricians, neonatologists) (stratum 2); policymakers (representing Ministry of Medical Services, World Health Organisation (WHO), United Nations Children's Fund (UNICEF)) (stratum 3); trainee paediatricians (stratum 4); and those with at least some experience of conducting systematic reviews (stratum 5).

By randomly allocating participants (in a 1:1:1 ratio within each of the 5 strata) to receive one of the three 'evidence pack + tracer-intervention' combination sets, we ensured: (1) that all participants received evidence on all three tracer-interventions; (2) that all participants were exposed to each of the three packaging formats; (3) that possible confounding of the relationship between evidence pack and the outcomes by tracer-intervention was reduced; (4) that subsequent interviewees would be able to reflect on the comparative value of each packaging format.

Figure 1 outlines the stratified randomisation process. One investigator (NO) generated the random allocation sequence (using a computer random number generator) and assigned participants to the different trial groups in sequential order. Participant recruitment was based on agreement to join the guideline development panel, attend its meeting and complete a self-administered questionnaire. We aimed to recruit a study population which represented those involved in guideline development and implementation in a low-income country.

Webappendix 3. Data handling and analysis

Primary outcome

a) Data handling

- The primary outcome was understanding of key information (as measured by the proportion of correct responses to key evidence pack findings).
- The 3-point responses (1=Correct, 2=Not clear, 3=Incorrect) were recoded into a binary response variable (0=Incorrect, 1=Correct).
- Each of the three tracer-interventions contributed two questions in the binary response variable (n=6 questions).
- All participants received each of the three tracer-interventions; hence, each participant contributed six responses in the overall binary response variable.

b) Method of analysis

- The odds of correct responses for pack B and C compared to the odds for pack A (assumed baseline pack) were estimated by calculating odds ratios (ORs) and 95% confidence interval (CIs) using logistic regression.
- To assess whether the effects of packs were modified by the type of strata, logistic regression model with type of pack and strata as an interaction term (pack*strata) was performed. Further tests of interactions were performed using likelihood ratio tests. Where evidence of interaction was found, stratum-specific ORs and 95% CIs were calculated.

Secondary outcome

a) Data handling

- The secondary outcome measure was mean 'value and accessibility' score. Value measure was participant self-report of perceived clarity of presentation of key information (1=Agree, 2=Not clear, 3=Disagree). Accessibility measures were self-reported ease of locating information on critical / important outcomes and judgments about the quality of evidence for critical / important outcomes; participants self-rated their responses on a 5-point scale.
- For both of the measures of access, we recoded the 5-point Likert scores (1=Strongly, 2=Disagree, 3=Neither agree nor disagree, 4=Strongly agree, 5=Agree) into 3-point Likert scores (1=Strongly disagree/Disagree, 2=Neither agree nor disagree, 3=Agree/Strongly agree).
- The 3-point 'value and accessibility' scores (responses) were further recoded to assume a common direction – meaning that scores ranged from 1 to 3 with the highest score interpreted as the most 'valuable / accessible'.
- Each of the three tracer-interventions contributed three questions to the 3-point Likert 'value and accessibility' score. Hence, each participant contributed 9 responses in the overall 'value and accessibility' score.

b) Method of analysis

- A mean 'value and accessibility' score for each of the three tracer-interventions was derived by summing up scores of responses (range, 3 to 9) and dividing by three. Thus, each individual contributed three mean scores to the analysis.
- The 'value and accessibility' scores were not normally distributed. (See histogram, figure 2). Methods that correct for non-normal outcome distributions were therefore applied (i.e. resampling by 'bootstrapping'[24]) to calculate estimates of pack effects.
- The mean 'value and accessibility' scores of pack B and pack C compared to the mean scores of pack A were estimated using linear regression.
- To confirm the results of linear regression analysis, an alternative approach to the analysis was undertaken: the odds of a one point increase in the 'clarity and accessibility' scores of pack B and C compared to the odds of pack A were estimated using ordinal logistic regression models. Further tests of interactions between pack and strata were performed using likelihood ratio tests. Where evidence of interaction was found, stratum-specific ORs and 95% CIs were calculated.

All analyses were done with STATA (version 11.0).

